# How to use CMVC with National Language Support (NLS) and Double-Byte Character Sets (DBCS)

Document Number TR 29.2169

Angel Rivera

CMVC/TeamConnection Development
IBM Software Solutions
Research Triangle Park, North Carolina
Copyright (C) 1996 IBM

# ABSTRACT

This technical report describes the overall background of using CMVC with National Language Support (NLS) and Double-Byte Character Sets (DBCS).

## ITIRC KEYWORDS

CMVC

NLS

DBCS

SBCS

# ABOUT THE AUTHORS

## ANGEL RIVERA

Mr. Rivera is a Staff Programmer and team lead for the CMVC development. He joined IBM in 1989 and since then has worked in the development and support of library systems.

Mr. Rivera has an M.S. in Electrical Engineering from The University of Texas at Austin, and a B.S. in Electronic Systems Engineering from the Instituto Tecnológico y de Estudios Superiores de Monterrey, México.

# CONTENTS

**ABSTRACT** . . . iii
ITIRC KEYWORDS . . . iii

**ABOUT THE AUTHORS** . . . v
Angel Rivera . . . v

**Figures** . . . viii

**Introduction** . . . 1

**Overall View** . . . 3

**General assumptions** . . . 5

**Relationships between NLSPATH, LANG and the CMVC message catalogs** . . . 7
Introduction . . . 7
Appropriate settings for each platform . . . 7
AIX 3.2.5 . . . 7
HP-UX 9 . . . 7
Solaris . . . 8
OS/2 . . . 8
DOS/Windows . . . 8
Using the Report command to test the NLS settings . . . 9

**DBCS Limitations** . . . 11

**How to install new locales and fonts in AIX** . . . 13
When using LANG=C, the message catalogs are ignored! . . . 13

**How to change NLS settings for CMVC line commands on HFT terminals** . . . 15

**How to change NLS settings for AIX CMVC GUI on aixterm windows** . . . 17

**Copyrights, Trademarks and Service marks** . . . 19

# FIGURES

# INTRODUCTION

This technical report describes the overall background on CMVC and National Language Support (NLS) and Double-Byte Character Sets (DBCS).

This technical report also provides instructions on how to use the CMVC client for AIX with different locales (which determine the code page) and their appropriate fonts. A common mistake with the AIX Client is to change only the locale but not the set the appropriate font that is associated with the code page for that locale.

All the steps mentioned in this note for the AIX client were successfully tested in an environment with AIX 2.3.5, with the locales En_US (IBM-850), en_US (ISO-8859-1) and an English keyboard. The database is DB2 or Oracle7, using En_US and Ja_JP locales. The server code is CMVC 2.3 (although CMVC 2.1 or 2.2 should work the same). Also, for DBCS we tested with a Japanese keyboard using the Ja_JP and the ja_JP locales.

The CMVC client for OS/2 was tested for DBCS on a PS/55, with OS.2 2.1J and a Japanese keyboard, using the 932 code page. Also we tested the CMVC client for general NLS issues from a PS/2 with OS/2 2.1 Spanish using a Spanish keyboard.

The CMVC client for Windows was tested for DBCS on a PS/2 Model 80 with DOS Version 5 Japanese, and Windows 3.1. The version of TCP/IP was 2.1.0.4. Also, we tested the code under the WIN-OS2 environment of OS/2 2.1J.

# OVERALL VIEW

The CMVC clients and the CMVC server do not alter the code points of the characters entered by the user, nor the characters exchanged with the database. Because most of the code pages have the same code points for the characters with value less than 128, then in practice there is no effect in using different code pages in this situation. However, if the customer wants to use characters with code points greater than 128 (such as accented characters, umlauts, etc.) then the code pages differ greatly in this respect. For example, the character with value 252 (which can be entered by pressing ALT and typing 2, 5 and 2 from the numeric keypad) has the representation of a lower case "u" with diaeresis (umlaut) when using the ISO-8859-1 code page, but has the representation of a superscript digit 3 when using the IBM-850 code page or the superscript digit n when using the IBM-437 code page.

In this cases, it is important that the customer should use the same code page in the CMVC client, in the CMVC server and in the database in order to avoid confusion if multiple code pages are used in these components.

The CMVC client for AIX Version 2.x can work with code pages that are ISO-8859 and non-ISO (such as IBM 850). By default, the code page installed with AIX 3.2 is the IBM-850 code page and the locale En_US. Thus, the Cmvc resource file that is used by the CMVC GUI has fonts defined that work only for the IBM-850 code page; also, for DBCS this resource file must be updated to use DBCS fonts.

The CMVC client for DOS/Windows uses the Windows code page which is practically the same as the ISO-8859-1 code page.

The CMVC client for OS/2 in the USA uses either the code page 437 or the code 850. For DBCS the code page is 942.

There is no support for Bi-Di languages in the CMVC clients.

# GENERAL ASSUMPTIONS

This technical report takes into account the following assumptions.

1. The keyboard to be used is the English keyboard, and thus, the extended characters are entered by means of pressing and holding the Alt key and then entering the 3 digits that correspond to the code point in decimal, using the numeric keypad. (Sorry, I do not understand why the numeric row does not work for this situation, only the numeric keypad in the right side of the keyboard works here). This action can be summarized as: Alt-XXX, such as Alt-252 which is pressing Alt, then entering the digits 2, followed by 5 and 2, then releasing the Alt key.

   Of course, if you have a keyboard that can enter extended characters, then you do not have to use the Alt-XXX method.

2. For testing purposes, some few characters are shown here with their code points for IBM-850 and ISO-8859-1:

| Character: | IBM-85 | ISO-8859-1 |
|---|---|---|
| u-umlaut | 129 | 252 |
| U-umlaut | 154 | 22 |
| e-umlaut | 137 | 235 |
| E-umlaut | 211 | 2 3 |
| o-umlaut | 148 | 246 |
| O-umlaut | 153 | 214 |
| ss (Beta) | 225 | 223 |
| a-grave-accent | 16 | 225 |
| A-grave-accent | 181 | 193 |

3. The quick test used in this note is to enter character 252. If the code page is IBM-850, then it will be a superscript "3" (that is, a tiny "3", raised), and if the code page is ISO-8859-1, then it will be the u-umlaut (the "u" with diaeresis, that is, with 2 dots on top).

4. The extended characters (or special characters) can be used in the server for configurable fields and in the items of the *.ld files such as config.ld; these characters can be used in the values for user names, description of components, etc...

5. For DBCS, the Japanese keyboard was used in a workstation for AIX, and in a workstation for OS/2 and DOS/Windows.

# RELATIONSHIPS BETWEEN NLSPATH, LANG AND THE CMVC MESSAGE CATALOGS

## INTRODUCTION

In the UNIX systems the variables LANG and NLSPATH are used to specify the locale and the location of the message catalogs respectively. The CMVC programs will use these variables to fetch the appropriate message from the CMVC message catalogs; if for some reason the message catalog cannot be found, then the CMVC programs will use the default English messages which are stored in the executables during the compilation process.

The exception is the SunOS platform which does not have the same facilities as in AIX, HP-UX and Solaris, with the consequence that the CMVC message catalogs are not used in SunOS, and that is the reason these message catalogs are not shipped for this platform.

See “When using LANG=C, the message catalogs are ignored!” on page 13 for more details when using the C locale.

## APPROPRIATE SETTINGS FOR EACH PLATFORM

### AIX 3.2.5

In AIX 3.2.5 the CMVC message catalogs are stored in:

```
/usr/lib/nls/msg/En_US
```

This means that the NLSPATH should have the strings:

```
export NLSPATH=/usr/lib/nls/msg/%L/%N
```

Where %L is the locale defined by LANG and %N is the actual name of the message catalog to be used. In this case, the value for the variable LANG should be:

```
export LANG=En_US
```

### HP-UX 9

In HP-UX 9 the CMVC message catalogs are stored in:

```
/usr/lib/nls/C
```

This means that the NLSPATH should have the strings:

```
export NLSPATH=/usr/lib/nls/%L/%N
```

Where %L is the locale defined by LANG and %N is the actual name of the message catalog to be used. In this case, the value for the variable LANG should be:

```
export LANG=C
```

## Solaris

In Solaris the CMVC message catalogs are stored in:

```
/usr/lib/locale/C/LC_MESSAGES
```

This means that the NLSPATH should have the strings:

```
export NLSPATH=/usr/lib/locale/%L/LC_MESSAGES/%N
```

Where %L is the locale defined by LANG and %N is the actual name of the message catalog to be used. In this case, the value for the variable LANG should be:

```
export LANG=C
```

## OS/2

In OS/2 the CMVC message catalog is stored in:

```
c:\cmvc\exe\nls
```

This means that the NLSPATH environment variable should have the strings:

```
set NLSPATH=c:\cmvc\exe\nls\%N
```

Where %N is the actual name of the message catalog to be used. Also, the NLS Path value in the Settings page should have this value. The %N must be in uppercase (if you use %n, then the message catalog will not be found).

## DOS/Windows

In DOS/Windows the CMVC message catalog is stored in:

```
c:\cmvcwin\exe\nls
```

This means that the NLSPATH variable in the Settings should have the string:

```
NLSPATH=c:\cmvcwin\exe\nls\%N
```

Where %N is the actual name of the message catalog to be used. The %N must be in uppercase (if you use %n, then the message catalog will not be found).

## USING THE REPORT COMMAND TO TEST THE NLS SETTINGS

You can use the command "Report -testServer" to verify if the CMVC server is accessing the message catalogs and the command "Report -testClient" to verify if the CMVC client is also accessing the message catalogs.

Some examples are shown below (the actual values might be different):

1) From OS/2:
```
c:\ Report -testServer
"The message catalog is available (1.285.2.118)"
```

2) From OS/2:
```
c:\ Report -testClient
"The message catalog is available (1.3)"
```

3) From AIX:
```
$ Report -testServer
"The message catalog is available (1.285.2.118)"
```

4) From AIX:
```
$ Report -testClient
"The message catalog is available (1.285.2.118)"
```

See “When using LANG=C, the message catalogs are ignored!” on page 13 for more details when using the C locale.

# DBCS LIMITATIONS

1. The administration tools for the CMVC Server expect SBCS characters as the reply for Yes (y) and No (n).
2. The administration tools for the CMVC Server have the following limitations for DBCS:

   a. The *.ld files (authority, interest, cfgcomproc and cfgrelproc) in the family account can accept DBCS characters in the first field for each entry. The maximum size for this field is 15 bytes.

   b. The config.ld file in the family account can accept DBCS characters in the following fields:

      Field 1: limit is 15 bytes

      Field 2: limit is 15 bytes

      Field 6: limit is 63 bytes

   c. The chfield program can accept only SBCS characters in the following fields:

      CMD attribute

      DB Column Name

   d. The chfield program can accept DBCS characters in the following fields:

      Field label: limit is 15 bytes

      Title label: limit is 15 bytes

      Type: must be a valid type defined in config.ld (limit is 15 bytes).

3. The CMVC User Manual, in Chapter 6 "CMVC Views for Queries" and in Chapter 7 "CMVC Tables for Queries", shows the datatype and the size limit for the attributes of the CMVC objects; however, the actual size limit for many of the character attributes is smaller than the specified limit. For example, the field "login" in the "Users" table shows that the limit is 31 characters, but in reality is only 15 bytes. This means that only 15 single byte characters can be stored, or 7 double byte characters. The fields affected are usually related to names, such as the User login, the Component name, etc.

   If you specify a string that has DBCS characters and that the size of the string goes beyond the limit, then the following error message will be displayed by the CMVC server:

1 -149 Your request cannot be completed.
The attribute flag argument xxx is not valid.
A frequent reason for this error is the presence of
characters that have a special meaning to the UNIX shell
such as $, double quotes, single quotes, back-slash.
If the string contains DBCS characters, then this message
may indicate that the size of the string is bigger than
the established limit.

Notice that the last sentence describes this situation, even though the beginning of the error message seems to have little relationship with the actual problem.

4. Warning on the use of 0x7C as a second byte in a DBCS character

   The Ox7C character corresponds to the vertical bar ('|') which in CMVC is interpreted as a field separator when dealing with reports and with handling windows and fields in the GUI.

   You can use this value as the 2nd byte of a DBCS character, however, when the data that contains this 2nd byte is handled in a CMVC client that has an SBCS code page (and not a DBCS code page), then, the output shown by the CMVC client may be displaced, that is, the 0x7C value will be interpreted as the field separator. Moreover, this situation will apply for any string in the *.ld files and in the configurable fields.

# HOW TO INSTALL NEW LOCALES AND FONTS IN AIX

Consult the Red Book "AIX Version 3.2 for RISC System/6000, National Language Support" (GG24-3850) for instructions on how to install locales and fonts.

For DBCS, it is important to install the Kanji fonts for X11.

Even though the /usr/lib/nls/msg directory contains subdirectories for other locales, it is not enough just to add the CMVC message catalogs into these subdirectories, because if the locale is not actually installed, these subdirectories will be ignored. You can use the command "locale -a" to find out which locales have been actually installed in your system.

## WHEN USING LANG=C, THE MESSAGE CATALOGS ARE IGNORED!

In the IBM red book "AIX 3.2 National Language Support", GG24-3850, says in page 16: "when the default C locale is the current locale, the hardcoded default message catalogs in the executable code will be used instead of the translated message file." The main point for confusion is that there is a directory in /usr/lib/nls/msg/C which is totally ignored! Thus, the user can place the message catalogs there, but they will never be used! This means that if the CMVC server is using the LANG=C locale, then the output for Report -testServer will always be that the message catalog is not available.

# HOW TO CHANGE NLS SETTINGS FOR CMVC LINE COMMANDS ON HFT TERMINALS

If you want to use CMVC without AIXwindows and work directly with an HFT terminal, you will need to setup the proper locale and the appropriate font. It is assumed at this point that you have installed the appropriate locales and fonts. The instructions to change the settings are:

1. logon as root
2. Type the following to invoke smit using the language menus:

   smit mlang
3. change language environment
4. select your language

   use F4 to select the choices, such as:

   English USA IBM-85 : En_US
   English USA ISO88591: en_US
5. manage fonts
6. list fonts in the system

   The list should have at least one, for example:

   Roman Medium IBM-85
7. If you want to work with an ISO code page, then you MUST install and setup an ISO font.
8. add a font to the system

   use F4 to select the pathname, such as: Rom22.iso1.snf
9. select the font palette
10. Primary Font IDENTIFIER
11. use F4 to select IBM-950 or ISO88591 font

    It is very likely that the system will display a warning saying that after the execution of the current command, smit will terminate.
12. exit from the root session
13. logon again as root
14. Test which Code Page are you using:

    a. Press and hold the Alt key

    b. Enter 3 digits in the numeric pad, such as: 2, 5 and 2

c. The previous 2 steps can be summarized as: Alt-252

d. If you are using IBM-850, then after typing Alt-252 you will see a tiny digit "3" (super-script).

e. If you are using ISO-8859-1, then after typing Alt-252 you will see the u-umlaut, that is, a lowercase "u" with diaeresis (2 dots) on top.

# HOW TO CHANGE NLS SETTINGS FOR AIX CMVC GUI ON AIXTERM WINDOWS

If you are going to use the CMVC line commands and the CMVC GUI in AIX, then you will be using the AIXwindows (Motif with X11).

1. Open an X-window session with aixterm with the desired locale (code-page) and the appropriate font. The -T "string" parameter for aixterm will determine the title for the window.

   For IBM-850:

   ```
   aixterm -T "IBM-85 " -lang En_US -fn Rom14 &
   ```

   For ISO-8859-1:

   ```
   aixterm -T "ISO-8859-1" -lang en_US -fn Rom14.iso1 &
   ```

   For Ja_JP:

   ```
   aixterm -T "Ja_JP" -lang Ja_JP -fn  gothic- -19  &
   ```

   For ja_JP:

   ```
   aixterm -T "ja_JP" -lang ja_JP -fn  gothic- -19  &
   ```

2. Perform the quick test for the code page, by entering Alt-252. If code page 850, then you will see 3-superscript; if code page ISO-8859-1, you will see u-umlaut.

   For DBCS, ensure that the aixterm window has the keyboard status area located at the bottom of the window; also, ensure that you can enter SBCS, Kana and Kanji DBCS characters.

3. Update the Cmvc resource file (the first C of this file name is in uppercase, and the rest of the string is in lowercase) to specify the appropriate font.

   The Cmvc file provided in the /usr/lpp/cmvc/samples directory is OK for IBM-850, and will not work for ISO code pages. However, in order to perform the tests indicated in this note, you will need to specify a different font and make 2 copies of this file:

   a. Copy /usr/lpp/cmvc/samples/Cmvc into Cmvc.850

   b. Change all lines that contain the string "fontList" to use the following font:

      ```
      Rom14
      ```

      For example, the first entry that has this string should be now:

      ```
      XmLabel fontList:                    Rom14
      ```

   c. Copy Cmvc.850 into Cmvc.iso1

   d. Change all instances of "Rom14" in Cmvc.iso1 to "Rom14.iso1".

e. Copy either Cmvc.850 or Cmvc.iso1 into Cmvc

f. For DBCS, you will need to specify the appropriate DBCS fonts:

1) Edit /usr/lpp/cmvc/samples/Cmvc

2) Change all lines that contain the string "fontList" to use the following font:

gothic- -19 :

For example, the first entry that has this string should be now:

XmLabel fontList: gothic- -19 :

**Important Note:** You must specify the colon at the end of the font specification; if the colon is not present, such as in:

gothic- -19

instead of:

gothic- -19 :

then the keyboard status area will not be shown.

4. Bring up the CMVC GUI by entering: cmvc
5. From the Windows pulldown, select "Users".
6. Create a user with at least the following data:

User ID: nls
User's full name: alt-252
Again, if using 85 , then you will see 3-superscript,
if using ISO, then you will see u-umlaut.
<Enter the rest of the data as you wish>

For DBCS: experiment entering SBCS and DBCS characters.

7. Query the Users window for user "nls". You should see an entry for this user, and in the 2nd column (Name) you should see either 3-superscript (850) or u-umlaut (ISO).

   For DBCS, you should see the appropriate SBCS/DBCS characters.

8. Double-click on that entry to show the CMVC Information window with the details. Ensure that the Name field has the appropriate character in it.
9. For DBCS, from the Users window select the Options pulldown. Then select "Set Family".

   You should see the keyboard status area at the bottom of the window and you should be able to enter SBCS and DBCS characters.

10. For DBCS, repeat step 8 above but this time use the File pulldown and select "Open List...".

    You should see the keyboard status area at the bottom of the window and you should be able to enter SBCS and DBCS characters.

# COPYRIGHTS, TRADEMARKS AND SERVICE MARKS

The following terms used in this technical report, are trademarks or service marks of the indicated companies:

| TRADEMARK, REGISTERED TRADEMARK OR SERVICE MARK | COMPANY |
|---|---|
| AIX, OS/2, IBM, DB2/6 , DB2, VisualGen, CMVC, ImagePlus, TeamConnection | IBM Corporation |
| UNIX, USL | UNIX System Laboratories, Inc. |
| NetLS, Network Licensing System | Apollo Computer, Inc., a subsidiary of Hewlett-Packard Co. |
| OSF, OSF Motif | Open Software Foundation, Inc. |
| PostScript | Adobe Systems Incorporated |
| INFORMIX | Informix Inc. |
| ORACLE | Oracle Corp. |
| SYBASE | Sybase Inc. |
| Sun, SunOS, OpenWindows, Solaris | Sun Microsystems Inc. |
| HP, HP-UX, SoftBench | Hewlett-Packard Company |
| Microsoft, Windows | Microsoft Corporation |
| X Window System | Massachusetts Institute of Technology |

**END OF DOCUMENT**