Configuration and Administration of the Unicode/Code Page Environment

This chapter covers the following topics:

- ICU Library

- Profile Parameters

- Encoding Information

- Deploying Natural Objects with Encoding Information

ICU Library

The ICU libraries are always installed with the full set of ICU conversion and collation data. The settings in the configuration file *NATCONV.INI* apply to the A format. For the U format, the corresponding checks (for example, when a character is translated to upper case) are done via the ICU library.

Note:
For obtaining information on the ICU version and the supported code pages, you can use the SYSCP utility which is available with Natural for Windows.

Profile Parameters

This section lists the profile parameters which are used in conjunction with Unicode and code page support.

Parameter	Description
CP	Defines the default code page for Natural. This code page is used for the runtime and development environment if not superposed with a code page defined for a single object (for example, for a Natural source). Only platform-suitable code pages can be used. This means, for example, that no EBCDIC code page can be defined for a Windows, UNIX or OpenVMS platform.
CPCVERR	Specifies whether a conversion error that occurs when converting from Unicode to code page or from code page to Unicode or from one code page to another code page results in a Natural error or not. This parameter is not regarded for the conversion of Natural sources when loading them into the source area or when cataloging them.
CPOBJIN	Specifies the code page in which the batch input file for data is encoded. This file is defined with the Natural profile parameter CMOBJIN .
CPPRINT	Specifies the code page in which the batch output file shall be encoded. This file is defined with the Natural profile parameter CMPRINT.
CPSYNIN	Specifies the code page in which the batch input file for commands is encoded. This file is defined with the Natural profile parameter CMSYNIN (Windows, UNIX and OpenVMS).
SRETAIN	Specifies that all existing sources have to be saved in their original encoding format. See also *Customizing Your Environment*.
SUTF8	Specifies the default format to be used when Natural sources are saved. **Note:** On UNIX and OpenVMS, this parameter can only be used in a SPoD environment.
SUBCHAR	Specifies the substitution character for the conversion from Unicode to the default code page. If SUBCHAR is OFF, the default substitution character defined by ICU will be used. **Note:** SUBCHAR does not influence conversions from code page to Unicode or from Unicode to code pages which differ from the default code page.
WEBIO	Specifies whether the Natural Web I/O Interface client (which supports Unicode) or the terminal emulation window (which is not Unicode-enabled) is used for input and output. In a local Windows environment, the output window (which is Unicode-enabled) is used. In a remote Windows environment, the Natural Web I/O Interface client is always used, regardless of the setting of this parameter.

See also:

- *Code Pages for the Input and Output Files* in the section *Natural in Batch Mode* of the *Operations* documentation

- For valid code pages, see *http://www.iana.org/assignments/character-sets*.

Encoding Information

The encoding of code page data can be specified on different levels.

Level 1 - Default Code Page

The default code page can be defined with the `CP` parameter. It overwrites the system code page and is valid for all code page data.

Level 2 - Code Page for a Single Object

A code page can be defined for Natural sources, batch input (`CPOBJIN`, `CPSYNIN`) and output files (`CPPRINT`).

In addition, a code page can be defined for work files of type ASCII, ASCII compressed, Unformatted and CSV; see *Work File Assignments* in the *Configuration Utility* documentation.

If a code page is defined at object level, this overwrites the default code page.

Important:
It is important that the correct code page is defined for every object. For more information, see *Migrating Existing Applications*.

Deploying Natural Objects with Encoding Information

If you want to deploy Natural objects for which encoding information has already been defined, you have to keep in mind that the encoding information is stored in the file *FILEDIR.SAG* and that it is lost if you deploy only the object file from outside of Natural.

When deploying Natural objects, you have the following possibilities for keeping the encoding information:

- You can copy the entire library. The copy of the library can then be distributed to all Windows, UNIX and OpenVMS platforms. In this case, the original code page is kept. If a library is copied from Windows to UNIX or OpenVMS, you have to keep in mind that it may be possible that the objects cannot be opened with a native Natural for UNIX or Natural for OpenVMS editor because these editors can only open objects with the default code page.

- You can use the Object Handler which keeps the encoding information. In this case, the original code page is kept. If a Windows library is unloaded on UNIX or OpenVMS, you have to keep in mind that it may be possible that the objects cannot be opened with a native Natural for UNIX or Natural for OpenVMS editor because these editors can only open objects with the default code page.

- You can copy and paste objects with Natural Studio. In a SPoD environment, if the target environment is located on a platform different from the source environment, Natural tries to save the object with the default code page of the target environment. If this is not possible, the object is stored in UTF-8 format. For UNIX and OpenVMS targets, this assures that the object can be opened with the native Natural for UNIX or Natural for OpenVMS editors, if all characters of the source are available in the default code page of the UNIX or OpenVMS server.