

Configuration and Administration of the Unicode/Code Page Environment

This chapter covers the following topics:

- ICU Library
 - Customizing the ICU Data Library for Mainframe Platforms
 - Profile Parameters
 - Encoding Information
 - Deploying Natural Objects with Encoding Information
-

ICU Library

Windows, UNIX and OpenVMS Platforms

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.

Mainframe Platforms

The relevant modules can be linked statically to the shared nucleus or loaded dynamically by means of the RCA technique. See *NATICU Modules for Different Purposes*.

For running applications without Unicode and without code page support, that is, with the profile parameter settings `CFICU=OFF` and `CP=OFF`, none of the supplied ICU modules needs to be linked to the Natural nucleus.

Note:

Information on the currently used ICU version and Unicode specification is provided in the main menu of the SYSCP utility. See *Invoking and Terminating SYSCP* in the *Utilities* documentation of the Natural for Mainframes documentation.

Three different load modules are offered:

Load Module	Description
NATICU	<p>Contains code page and Unicode conversion functionality as well as collation services. The first is needed for conversion from one code page to another code page or to Unicode and vice versa. The latter is used for string comparison of Unicode strings with consideration of locale ID.</p> <p>NATICU contains the most popular code pages and locales. The code pages are already declared in NATCONFIG.</p>
NATICUCV	<p>Same as NATICU, but without collation services. Therefore, it is smaller.</p>
NATICUXL	<p>Same as NATICU, but it contains all possible converters and locales offered by the currently supported ICU version. It supports about 230 different code pages (predominantly EBCDIC code pages) and 238 locales. Therefore, the module size is huge.</p> <p>If NATICUXL is linked to the Natural nucleus, the desired code pages have to be declared in the configuration file NATCONFIG.</p> <p>NATICUXL supports all code pages and locale IDs which are supported by the currently supported ICU version (see http://demo.icu-project.org/icu-bin/convexp).</p> <p>Note: Due to technical restrictions, NATICUXL is not delivered for z/VSE.</p>

See also *NATICU Modules for Different Purposes*.

Note for z/VSE:

If you receive an error during linkage editing when you try to link one of the ICU load modules statically to the shared nucleus because the size of the resulting phase is too large for your z/VSE partition, proceed as follows:

- Check whether NATICUCV can be used instead of NATICU (if you tried to link NATICU).
- Load the relevant module dynamically by means of the RCA technique. See *NATICU Modules for Different Purposes*. Modules NATICU and NATICUCV are also delivered as a phase to make use of the RCA technique more convenient.
- Ask your system administrator to configure your partition so that the storage available for the linkage editor is sufficient to hold the resulting phase.

Additional tables:

Table	Description
NATSCTU	Required scanner table for Unicode characters. It maps the properties of Unicode characters of the currently supported Unicode version to be used by the Natural nucleus. This table must never be changed.
NATCPTAB	Optional single-byte code page conversion accelerator tables. If the table is present, conversion from one code page to another code page will be faster since it is performed via this table rather than by calling ICU functions. For more information, see <i>Translation Tables</i> in the <i>Operations</i> documentation for Natural for Mainframes.

Customizing the ICU Data Library for Mainframe Platforms

ICU makes use of a wide variety of data tables to provide many of its services. Examples include converter mapping tables, collation rules, transliteration rules, break iterator rules and dictionaries, and other locale data. The ICU data library for Natural is provided as a package that contains the desired data items. The usage of packages instead of single data item files increases the performance since there is only one file access during the initialization to load the package. However, it is not so flexible since it requires a rebuild of a package if data items need to be added.

The ICU data library may be customized in order to add existing or new converter mapping tables or to add other data items such as collation rules, break iterator rules and other locale data.

The customization tool for the ICU data library is available from the Download Components area in Empower (<https://empower.softwareag.com/>). Detailed information on how to customize the ICU data library is provided in the *readme.txt* file which is part of the downloaded zip file.

Profile Parameters

This section lists the profile parameters which are used in conjunction with Unicode and code page support. Not all profile parameters are available on all platforms.

- All Platforms
- Windows, UNIX and OpenVMS Platforms
- Mainframe Platforms
- Natural Development Server for Mainframes

All Platforms

The following profile parameters are available on all platforms:

Parameter	Description
CP	<p>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).</p> <p>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, or that no ASCII code page can be defined for a mainframe platform. On mainframes, an initialization error message occurs if a wrong code page is used.</p> <p>Note: As of Natural Version 6.2 for Windows and UNIX, Natural Version 6.3 for OpenVMS and Natural Version 4.2 for Mainframes, the existing CP parameter (used with Natural RPC) has been renamed to CPRPC.</p>
CPCVERR	<p>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.</p> <p>This parameter is not regarded for the conversion of Natural sources when loading them into the source area or when cataloging them.</p> <p>On mainframe platforms, it is not regarded whether a Unicode field is converted into the code page before an I/O on a terminal emulation. In this case, the substitution character defined by ICU is replaced by the place holder character which is defined in NATCONFIG.</p>
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 (Windows, UNIX and OpenVMS) or in the dataset CMOBJIN (mainframe).
CPPRINT	Specifies the code page in which the batch output file shall be encoded. This file is defined with the Natural profile parameter CMPRINT (Windows, UNIX and OpenVMS) or in the dataset CMPRINT (mainframe).
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) or in the dataset CMSYNIN (mainframe).
SRETAIN	Specifies that all existing sources have to be saved in their original encoding format. See also <i>Customizing Your Environment</i> .

See also:

- *Code Pages for the Input and Output Files* in the section *Natural in Batch Mode* of the *Operations* documentation for Natural for Windows, Natural for UNIX and Natural for OpenVMS.
- *Natural in Batch Mode* in the *Operations* documentation for Natural for Mainframes.
- For valid code pages, see <http://www.iana.org/assignments/character-sets>.

Windows, UNIX and OpenVMS Platforms

The following profile parameters are only available on Windows, UNIX and OpenVMS platforms:

Parameter	Description
SUTF8	<p>Specifies the default format to be used when Natural sources are saved.</p> <p>Note: On UNIX and OpenVMS, this parameter can only be used in a SPoD environment.</p>
SUBCHAR	<p>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.</p> <p>Note: SUBCHAR does not influence conversions from code page to Unicode or from Unicode to code pages which differ from the default code page.</p>
WEBIO	<p>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.</p> <p>In a local Windows environment, the output window (which is Unicode-enabled) is used.</p> <p>In a remote Windows environment, the Natural Web I/O Interface client is always used, regardless of the setting of this parameter.</p> <p>Note: For mainframe platforms, the NDV configuration parameter <code>TERMINAL_EMULATION</code> is used instead. See below.</p>

Mainframe Platforms

The following profile parameters and/or macros are only available on mainframe platforms:

Parameter	Macro	Description
CFICU	NTCFICU	Enables Unicode support for various Unicode settings.
Not available.	NTCPAGE	In the NATCONFIG module, this macro defines a code page and all related information, such as replacement characters, locale ID and collation tables.
PRINT	CP keyword subparameter of NTPRINT macro	Defines the code page for a report.
CMPO	CPAGE keyword subparameter of NTCMPO macro	Generates code page-sensitive Natural programs.
OPRB	NTOPRB	Set the ACODE and/or WCODE option to define the user encoding if the used Adabas database is enabled for UES (universal encoding support).

Natural Development Server for Mainframes

The following NDV configuration parameter is only available with the Natural Development Server for mainframe platforms:

Settings	Description
TERMINAL_EMULATION=WEBIO	Specifies that the Natural Web I/O Interface client (which supports Unicode) is used for input and output.

Encoding Information

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

- Level 1 - Default Code Page
- Level 2 - Code Page for a Single Object

Level 1 - Default Code Page

The default code page can be defined with the CP parameter. On Windows, UNIX and OpenVMS platforms, 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 (CPRINT).

In addition, on Windows, UNIX and OpenVMS platforms, 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.

Note:

On Windows, UNIX and OpenVMS platforms, 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

Windows, UNIX and OpenVMS Platforms

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.

Mainframe Platforms

For objects on mainframe platforms, there are no special considerations for keeping the code page information of the object since it is part of the object directory.