

Locale String Mapping

It is assumed that you have read the chapter *Introduction to Internationalization* and are familiar with the various internationalization approaches described there.

A locale provides a means of identifying a specific region for the purposes of internationalization and localization. EntireX sends properties of the operating system locale to the Broker, using the locale string.

This chapter describes the mapping of the locale string to codepages within the broker for the internationalization approaches ICU conversion and SAGTRPC user exit. It does not apply to other approaches.

This chapter covers the following topics:

- Broker's Locale String Processing
- Broker's Built-in Locale String Mapping
- Broker's Locale String Defaults
- Configuring Broker's Locale String Defaults
- Bypassing Broker's Built-in Locale String Mapping
- Using the Abstract Codepage Name LOCAL

See also *Broker Attributes*.

Broker's Locale String Processing

Depending on the internationalization approach in effect for a service or topic, the result of the locale string processing within the broker is either the ICU converter alias or the codepage number given to SAGTRPC user exit.

There are always two EntireX components involved - client or server, sender or receiver - so for genuine conversion this process is run through for the sender to determine its codepage as well as for the receiver to get the codepage. It is important to know both codepages in order to predict conversion behavior accurately.

1. If no locale string is provided by an EntireX component (sender or receiver), the *Broker's Locale String Defaults* will apply. These can be customized in the broker attribute file; see *Configuring Broker's Locale String Defaults*.
2. If a locale string is provided by an EntireX component (sender or receiver), the broker first refers to the codepage section of the attribute file searching for a keyword entry identical to the locale string sent. You can also bypass these entries to fit your needs; see *Bypassing Broker's Built-in Locale String Mapping*.
3. If an entry is found in step 2, this codepage is used directly and no further mapping occurs.

4. If no entry is found in step 2, the *Broker's Built-in Locale String Mapping* is entered.

A *prerequisite* for the broker to use internationalization correctly is a suitably configured environment. This includes the broker's platform and may include the platforms that other EntireX components (sender and receiver) run on.

- The data that EntireX components send to the broker *must* be in the encoding described by the locale string.
- The data that EntireX components send to the broker *must* be in the encoding of the codepage the broker uses for conversion or translation.
- After broker's locale string processing (steps 1-4 above), the resulting ICU converter or the code points implemented with the Translation user exit or with the SAGTRPC user exit *must* match the defined code points of the original codepage of your application's environment. (Matching code points is not a trivial matter. Almost every hardware and software vendor provides its own codepage, based on standards organizations such as ISO etc. It is even more difficult to find matching codepages because there is no unique scheme of identifiers for codepages across all organizations.)

If one of the prerequisites above is not met, results will be unpredictable.

Broker's Built-in Locale String Mapping

The table in this section describes the built-in mechanism the broker uses to map a locale string to a codepage.

Whenever possible, we recommend using the abstract codepage name "LOCAL". Without adapting your EntireX component, this setting allows you to

- reconfigure your system codepage
- install your EntireX application in systems configured for different countries and regions using different codepages

See *Using the Abstract Codepage Name LOCAL*.

The form ECS *<ecs-number>* is deprecated and should no longer be used.

The last 2 forms in the table below CP *<number>* and *<codepage-name>* are the forms administrators and programmers use to configure or provide a codepage manually. This is necessary if

- a locale string is not sent by default to the broker, and
- the abstract codepage name LOCAL cannot be used.

Depending on the format of the locale string sent by the EntireX component, various rules for mapping apply:

Locale String Sent to Broker	Environment	Locale String Format Sent	ICU Converter Alias	SAGTRPC User Exit Codepage Number
<p><language>_<country>.<number> or <language>-<country>.<number></p> <p>where any abbreviation is accepted for <language> and <country> and <number> is the decimal number of a Windows codepage.</p>	Windows	<ul style="list-style-type: none"> by default if communicating with broker version 7.2.x and above if communicating with broker version 7.1.x and below and the <i>Using the Abstract Codepage Name LOCAL</i> is given. if manually configured or provided by programmer. 	<p>The part <number> is extracted from the locale string and prefixed by the term windows. The result is used as the ICU converter alias. The parts <language> and <country> are not used.</p> <p>Example: german_Germany.1252 results in the alias windows-1252 for the ICU converter alias.</p>	<p>The part <number> is extracted and used together with the table <i>Mapping Codepage Numbers</i> to evaluate the number given to the SAGTRPC user exit. The parts <language> and <country> are not used.</p>
<p><ll>_<cc> where <ll> is the 2-letter language abbreviation and <cc> the 2 or 3-letter language code according to ISO 639 standard.</p>	UNIX	<ul style="list-style-type: none"> if the <i>Using the Abstract Codepage Name LOCAL</i> is given. 	<p>The part <ll> is extracted and used together with the table <i>Mapping Two-character Language Codes</i> to evaluate the ECS codepage number. The part <cc> is not used.</p>	<p>The part <ll> is extracted and used together with the table <i>Mapping Two-character Language Codes</i> to evaluate the number given to the SAGTRPC user exit. The part <cc> is not used.</p>
<p><ll>_<cc>.<UNIX-codepage-name> where <ll> is the 2-letter language abbreviation, <cc> the 2 or 3-letter language code according to ISO 639 standard and <UNIX-codepage-name> is any alias name for a codepage.</p>	UNIX	<ul style="list-style-type: none"> if the <i>Using the Abstract Codepage Name LOCAL</i> is given. 	<p>The part <UNIX-codepage-name> is extracted from the locale string and used for the ICU converter alias. The parts <ll> and <cc> are not used.</p>	<p>The part <UNIX-codepage-name> is extracted and used together with the table <i>Mapping UNIX Codepage Names</i> to evaluate the number given to the SAGTRPC user exit. The parts <ll> and <cc> are not used.</p>

Locale String Sent to Broker	Environment	Locale String Format Sent	ICU Converter Alias	SAGTRPC User Exit Codepage Number
CP <number> where <number> is the decimal number of a codepage.	All	<ul style="list-style-type: none"> by default in Java environments if communicating with broker version 7.2.x and above if manually configured or provided by a programmer 	The locale string CP <number> is used as is for the ICU Converter alias.	The part <number> is extracted and used together with the table <i>Mapping Codepage Numbers</i> to evaluate the number given to the SAGTRPC user exit.
<codepage-name> where <codepage-name> is any alias name for a codepage.	All	<ul style="list-style-type: none"> by default in Java environments if communicating with broker version 7.2.x and above if manually configured or provided by a programmer 	The locale string <codepage-name>e is used as is for the ICU Converter alias.	The locale string <codepage-name> is used together with the table <i>Mapping Java Codepage Names</i> to evaluate the number given to the SAGTRPC user exit.

With ICU Conversion

- Once you have determined the ICU converter alias, use the ICU Converter Explorer under *ICU Resources* to determine the real ICU converter's canonical name and get more information on the ICU converter.

With SAGTRPC User Exit

- Once you have determined the number given to SAGTRPC user exit, the implementation of the codepage is your responsibility. No more information is provided.

Broker's Locale String Defaults

If a locale string is not sent by an EntireX component (sender or receiver), the broker itself makes a rough assumption to assign an *ICU Resources* or a numeric codepage number with SAGTRPC user exit.

The broker can distinguish between ASCII environments, IBM mainframe and Fujitsu mainframe operating systems if the EntireX component does not indicate anything by the locale string. See the following table:

Platform EntireX Component is running on	ICU Converter	SAGTRPC User Exit
UNIX, Windows	ibm-5349_P100-1998	1252
IBM Mainframe	ibm-37_P100-1995	0037
Fujitsu Mainframe	bs2000-edf04drv	3587

Note that the broker's built-in defaults above may match for the following countries:

- Many western countries using Windows configured with the ICU converter `ibm-5349_P100-1998` (an ICU-supported alias name is, for example, `CP1252`) including Java environments.
- The United States using OS/390 or z/OS IBM mainframe configured with the ICU converter `ibm-37_P100-1995` (ICU-supported alias names are `CP37` and `ibm-37`).

The defaults do not match in the following scenarios, and possibly other scenarios also:

- if codepages other than the ones listed above are used for an operating system
- for UNIX operating systems, including Java environments

You can customize the defaults to your own requirements. See the respective attribute in *Codepage-specific Attributes* under *Broker Attributes* for how to customize the broker's locale string defaults. For examples of how to configure the broker's locale strings, see next section, *Configuring Broker's Locale String Defaults*.

Configuring Broker's Locale String Defaults

The broker's built-in defaults for locale strings can be overridden by assigning the ICU converter name directly or an alias of the ICU converter. See *Codepage-specific Attributes* under *Broker Attributes*.

This procedure is useful

- if the built-in default does not meet your requirements
- if EntireX components (sender or receiver) do not send locale strings.

Example 1

For this example it is assumed that the internationalization approach is ICU conversion.

An environment running in Spanish-speaking countries using clients with Windows 1252 codepages and servers on IBM mainframe with codepage 1145. Because 1252 is the broker's default for ASCII environments, the default for IBM mainframe is changed to codepage 1145 only, using the ICU converter alias `ibm-1145`. The previously used codepage 284 is also possible, but does not contain the euro sign.

```
DEFAULTS=CODEPAGE
/* Broker Locale String defaults */
DEFAULT_EBCDIC_IBM=ibm-1145
```

As a result, the related ICU converter used as the default for IBM mainframe is `ibm-1145_P100-1997`. See ICU Converter under *ICU Resources*.

Example 2

For this example it is assumed that the internationalization approach is ICU conversion.

An environment running in German-speaking countries using Windows 1252 codepages and servers on IBM mainframe with codepage 1141. Because 1252 is the broker's default for ASCII environments, the default for IBM mainframe is changed to codepage 1141 only, using the ICU converter alias `ibm-1141`. The previously used codepage 273 is still possible but does not contain the euro sign.

```
DEFAULTS=CODEPAGE
/* Broker Locale String defaults */
DEFAULT_EBCDIC_IBM=ibm-1141
```

As a result, the related ICU converter used as the default for IBM mainframe is `ibm-1141_P100-1997`, see *ICU Converter Explorer* under *ICU Resources*.

Example 3

For this example it is assumed that the internationalization approach is ICU conversion.

An environment running in Hong Kong using clients with the Windows 950 (big5) codepage and servers on IBM mainframe with codepage 937. For suitable default values, the broker's default for ASCII environments as well as for IBM mainframes is adapted by assigning ICU converters' alias names.

```
DEFAULTS=CODEPAGE
/* Broker Locale String defaults */
DEFAULT_ASCII=windows-950
DEFAULT_EBCDIC_IBM=ibm-937
```

As a result, the related ICU converter used as the default is

- for ASCII environments: `ibm-1373_P100-2002`
- for IBM mainframe: `ibm-937_P110-1999`

See *ICU Converter Explorer* under *ICU Resources*.

Example 4

For this example it is assumed that the internationalization approach is ICU conversion.

An environment running in Turkey using clients with the Windows 1254 codepage and servers on IBM mainframe with codepage 1026. For suitable default values, the broker's default for ASCII environments as well as for IBM mainframes is adapted by assigning ICU converters' alias names.

```
DEFAULTS=CODEPAGE
/* Broker Locale String defaults */
DEFAULT_ASCII=windows-1254
DEFAULT_EBCDIC_IBM=ibm-1026
```

As a result, the related ICU converter used as the default is

- for ASCII environments: `ibm-1254_P100-1995`

- for IBM mainframe: ibm-1026_P100-1995

See *ICU Converter Explorer* under *ICU Resources*.

Bypassing Broker's Built-in Locale String Mapping

The broker's built-in mechanism of mapping locale strings to codepages can be bypassed by assigning the ICU converter name directly or an alias of the ICU converter. See *Codepage-specific Attributes* under *Broker Attributes*. Locale string matching is case-insensitive when bypassing the broker's built-in mechanism, that is, when the broker examines the DEFAULTS=CODEPAGE section in the attribute file.

- If an EntireX component (sender and receiver) sends a locale string where the broker's built-in mechanism fails and no codepage is found, you can assign a codepage.
- If an EntireX component sends a locale string where the broker's built-in mechanism selects the wrong codepage, you can assign the correct codepage.
- If you cannot adapt the locale string sent by your EntireX component when an incorrect locale string is sent.

Example

For this example it is assumed that the internationalization approach is ICU conversion.

An EntireX Java component sends ASCII as the locale string. This is mapped by ICU to US-ASCII; instead, ISO 8859_1 should be used. This is done with the following configuration:

```
DEFAULTS=CODEPAGE
/* Broker Locale String Codepage Assignments */
ASCII=ISO8859_1
```

Using the Abstract Codepage Name LOCAL

In *Windows environments*, the default Windows ANSI codepage configured for your system can also be determined automatically and sent to the broker if the abstract codepage name LOCAL as the locale string is given by the EntireX component. This requires configuration of the EntireX component for an administrator or coding it for a developer and is necessary if communicating with broker version 7.1.x and below. See *Preparing EntireX Components for Internationalization*. If communicating with broker version 7.2.x and above, the default Windows ANSI codepage is sent to the broker by default.

In *UNIX environments*, no locale string is sent to the broker by default. EntireX components must always be set up for internationalization. See *Preparing EntireX Components for Internationalization*. Using the abstract codepage name LOCAL requires a well-configured UNIX system for internationalization. See the environment variables LC_TYPE and LC_LANG and see your UNIX documentation.

In *Java environments* (and with most EntireX components), the default encoding configured for your Java virtual machine can also be determined automatically and sent to the broker if the abstract codepage name LOCAL as the locale string is given by the EntireX component. This requires configuration of the EntireX component for an administrator or coding it for a developer and is necessary if communicating with broker version 7.1.x and below. See *Preparing EntireX Components for Internationalization*. If communicating with broker version 7.2.x and above, the default encoding configured for your Java virtual machine is sent to the broker by default.

In *all other* operating systems and environments, the abstract codepage name LOCAL is *not* supported. See *Preparing EntireX Components for Internationalization*.