

# How Transliteration Works in matchIT

## Field based transliteration

With a compiled list of words for commonly used address field values in each language, data is transliterated or translated depending on the word.

## Character Mapping

Both native and Latin character sets are mapped for character based transliteration.



## Native to Latin character set

When an address is entered in a native language, but requires an English output, the address fields are queried for the transliterated value. If the address fields are not found, the individual characters are transliterated. If both attempts at transliteration do not work, the native word is returned.

Below is a list of supported character sets, along with the respective ISO-15924 codes. The default value is "Native" which means the output script will match the input wherever possible.

Language	Region	ISO-15924
Latin		LATN
Cyrillic	Russia	CYRL
Hellenic	Greece	GREK
Hebrew	Israel	HEBR
Kanji	Japan	HANI
Simplified Chinese	China	HANS
Arabic	United Arab Emirates	ARAB
Thai	Thailand	THAI
Hangul	South Korea	HANG



## Cleaner Data. Better Decisions.

For the past 20 years, helpIT systems has been tightly focused on developing and delivering data quality technology that generates tangible and accurate results. With over 2,000 clients in 30 countries across 5 continents, helpIT is consistently raising the bar on data quality success.

**Americas, Australia, New Zealand:** 866.332.7132

**UK, Europe, Asia:** 011 +44 (0) 1372 225 900

[www.helpIT.com](http://www.helpIT.com)