# NON ASCII Character translation

Significant bulk data files (for example AIIR, PHIR) were being rejected due to the detection of non-ASCII characters during the BDE validation process in the ATO portals.

In most cases however, the data was not of poor quality as validation has failed on a real name that happens to contain special characters (for example).

 In an effort to reduce the number of failed lodgements and delays in getting data to pre-fill, the way that BDE detected and processed non-ASCII characters for all file types needed to change.

**As BDE and ICP cannot process non-ASCIIs characters, a find and replace solution was deployed into Production in December 2015.**

The updated process:

* Once the file has been accepted by BDE - prior to the validation process - the translation and conversion process runs.
* This non-ASCII character pre-check runs against a defined list of special characters.
* If a conversion is required, a notification of this is communicated back to user as a warning message in the validation report that a specific record has be changed due to the detection of a non-ASCII character.

## Conversion table

Suggest conversion is only those with a value in ‘Convert to character’ column

| Character | Convert to Character | Comment |
| --- | --- | --- |
| ASCII code 129 = **ü** | U |  |
| ASCII code 128 = Ç ( Majuscule C-cedilla ) | **C** |  |
| ASCII code 129 = ü ( letter u with umlaut or diaeresis , u-umlaut ) | U |  |
| ASCII code 130 = é ( letter e with acute accent or e-acute ) | E |  |
| ASCII code 131 = â ( letter a with circumflex accent or a-circumflex ) | A |  |
| ASCII code 132 = ä ( letter a with umlaut or diaeresis , a-umlaut ) | A |  |
| ASCII code 133 = à ( letter a with grave accent ) | A |  |
| ASCII code 134 = å ( letter a with a ring ) | **A** |  |
| ASCII code 135 = ç ( Minuscule c-cedilla ) | **C** |  |
| ASCII code 136 = ê ( letter e with circumflex accent or e-circumflex ) | E |  |
| ASCII code 137 = ë ( letter e with umlaut or diaeresis ; e-umlauts ) | E |  |
| ASCII code 138 = è ( letter e with grave accent ) | E |  |
| ASCII code 139 = ï ( letter i with umlaut or diaeresis ; i-umlaut ) | **I** |  |
| ASCII code 140 = î ( letter i with circumflex accent or i-circumflex ) | **I** |  |
| ASCII code 141 = ì ( letter i with grave accent ) | **I** |  |
| ASCII code 142 = Ä ( letter A with umlaut or diaeresis ; A-umlaut ) | A |  |
| ASCII code 143 = Å ( Capital letter A with a ring ) | A |  |
| ASCII code 144 = É ( Capital letter E with acute accent or E-acute ) | E |  |
| ASCII code 145 = æ ( Latin diphthong ae in lowercase ) | **0** |  |
| ASCII code 146 = Æ ( Latin diphthong AE in uppercase ) | **0** |  |
| ASCII code 147 = ô ( letter o with circumflex accent or o-circumflex ) | **O** |  |
| ASCII code 148 = ö ( letter o with umlaut or diaeresis ; o-umlaut ) | **O** |  |
| ASCII code 149 = ò ( letter o with grave accent ) | **O** |  |
| ASCII code 150 = û ( letter u with circumflex accent or u-circumflex ) | u |  |
| ASCII code 151 = ù ( letter u with grave accent ) | U |  |
| ASCII code 152 = ÿ ( Lowercase letter y with diaeresis ) | Y |  |
| ASCII code 153 = Ö ( Letter O with umlaut or diaeresis ; O-umlaut ) | O |  |
| ASCII code 154 = Ü ( Letter U with umlaut or diaeresis ; U-umlaut ) | U |  |
| ASCII code 155 = ø ( Lowercase slashed zero or empty set ) | **O** |  |
| ASCII code 156 = £ ( Pound sign ; symbol for the pound sterling ) | **0** |  |
| ASCII code 157 = Ø ( Uppercase slashed zero or empty set ) | **O** |  |
| ASCII code 158 = × ( Multiplication sign ) | **X** |  |
| ASCII code 159 = ƒ ( Function sign ; f with hook sign ; florin sign ) | **F** |  |
| ASCII code 160 = á ( Lowercase letter a with acute accent or a-acute ) | A |  |
| ASCII code 161 = í ( Lowercase letter i with acute accent or i-acute ) | I |  |
| ASCII code 162 = ó ( Lowercase letter o with acute accent or o-acute ) | O |  |
| ASCII code 163 = ú ( Lowercase letter u with acute accent or u-acute ) | U |  |
| ASCII code 164 = ñ ( eñe, enie, spanish letter enye, lowercase n with tilde ) | **N** |  |
| ASCII code 165 = Ñ ( Spanish letter enye, uppercase N with tilde, EÑE, enie ) | N |  |
| ASCII code 166 = ª ( feminine ordinal indicator ) | **A** |  |
| ASCII code 167 = º ( masculine ordinal indicator ) | **O** |  |
| ASCII code 168 = ¿ ( Inverted question marks ) | **0** |  |
| ASCII code 169 = ® ( Registered trademark symbol ) | **0** |  |
| ASCII code 170 = ¬ ( Logical negation symbol ) | **0** |  |
| ASCII code 171 = ½ ( One half ) | **0** |  |
| ASCII code 172 = ¼ ( Quarter, one fourth ) | **0** |  |
| ASCII code 173 = ¡ ( Inverted exclamation marks ) | **0** |  |
| ASCII code 174 = « ( Angle quotes, guillemets, right-pointing quotation mark ) | **0** |  |
| ASCII code 175 = » ( Guillemets, angle quotes, left-pointing quotation marks ) | **0** |  |
| ASCII code 176 = ░ ( Graphic character, low density dotted ) | **0** |  |
| ASCII code 177 = ▒ ( Graphic character, medium density dotted ) | **0** |  |
| ASCII code 178 = ▓ ( Graphic character, high density dotted ) | **0** |  |
| ASCII code 179 = │ ( Box drawing character single vertical line ) | **0** |  |
| ASCII code 180 = ┤ ( Box drawing character single vertical and left line ) | **0** |  |
| ASCII code 181 = Á ( Capital letter A with acute accent or A-acute ) | A |  |
| ASCII code 182 = Â ( Letter A with circumflex accent or A-circumflex ) | A |  |
| ASCII code 183 = À ( Letter A with grave accent ) | A |  |
| ASCII code 184 = © ( Copyright symbol ) | **0** |  |
| ASCII code 185 = ╣ ( Box drawing character double line vertical and left ) | **0** |  |
| ASCII code 186 = ║ ( Box drawing character double vertical line ) | **0** |  |
| ASCII code 187 = ╗ ( Box drawing character double line upper right corner ) | **0** |  |
| ASCII code 188 = ╝ ( Box drawing character double line lower right corner ) | **0** |  |
| ASCII code 189 = ¢ ( Cent symbol ) | **C** |  |
| ASCII code 190 = ¥ ( YEN and YUAN sign ) | **0** |  |
| ASCII code 191 = ┐ ( Box drawing character single line upper right corner ) | **0** |  |
| ASCII code 192 = └ ( Box drawing character single line lower left corner ) | **0** |  |
| ASCII code 193 = ┴ ( Box drawing character single line horizontal and up ) | **0** |  |
| ASCII code 194 = ┬ ( Box drawing character single line horizontal down ) | **0** |  |
| ASCII code 195 = ├ ( Box drawing character single line vertical and right ) | **0** |  |
| ASCII code 196 = ─ ( Box drawing character single horizontal line ) | **0** |  |
| ASCII code 197 = ┼ ( Box drawing character single line horizontal vertical ) | **0** |  |
| ASCII code 198 = ã ( Lowercase letter a with tilde or a-tilde ) | A |  |
| ASCII code 199 = Ã ( Capital letter A with tilde or A-tilde ) | **A** |  |
| ASCII code 200 = ╚ ( Box drawing character double line lower left corner ) | **0** |  |
| ASCII code 201 = ╔ ( Box drawing character double line upper left corner ) | **0** |  |
| ASCII code 202 = ╩ ( Box drawing character double line horizontal and up ) | **0** |  |
| ASCII code 203 = ╦ ( Box drawing character double line horizontal down ) | **0** |  |
| ASCII code 204 = ╠ ( Box drawing character double line vertical and right ) | **0** |  |
| ASCII code 205 = ═ ( Box drawing character double horizontal line ) | **0** |  |
| ASCII code 206 = ╬ ( Box drawing character double line horizontal vertical ) | **0** |  |
| ASCII code 207 = ¤ ( Generic currency sign ) | **0** |  |
| ASCII code 208 = ð ( Lowercase letter eth ) | **O** |  |
| ASCII code 209 = Ð ( Capital letter Eth ) | **D** |  |
| ASCII code 210 = Ê ( Letter E with circumflex accent or E-circumflex ) | E |  |
| ASCII code 211 = Ë ( Letter E with umlaut or diaeresis, E-umlaut ) | E |  |
| ASCII code 212 = È ( Capital letter E with grave accent ) | E |  |
| ASCII code 213 = ı ( Lowercase dot less i ) | **I** |  |
| ASCII code 214 = Í ( Capital letter I with acute accent or I-acute ) | I |  |
| ASCII code 215 = Î ( Letter I with circumflex accent or I-circumflex ) | I |  |
| ASCII code 216 = Ï ( Letter I with umlaut or diaeresis ; I-umlaut ) | I |  |
| ASCII code 217 = ┘ ( Box drawing character single line lower right corner ) | **0** |  |
| ASCII code 218 = ┌ ( Box drawing character single line upper left corner ) | **0** |  |
| ASCII code 219 = █ ( Block, graphic character ) | **0** |  |
| ASCII code 220 = ▄ ( Bottom half block ) | **0** |  |
| ASCII code 221 = ¦ ( Vertical broken bar ) | | | Normal broken Bar ASCII 124 |
| ASCII code 222 = Ì ( Capital letter I with grave accent ) | I |  |
| ASCII code 223 = ▀ ( Top half block ) | **0** |  |
| ASCII code 224 = Ó ( Capital letter O with acute accent or O-acute ) | O |  |
| ASCII code 225 = ß ( Letter Eszett ; scharfes S or sharp S ) | **S** |  |
| ASCII code 226 = Ô ( Letter O with circumflex accent or O-circumflex ) | O |  |
| ASCII code 227 = Ò ( Capital letter O with grave accent ) | O |  |
| ASCII code 228 = õ ( Lowercase letter o with tilde or o-tilde ) | O |  |
| ASCII code 229 = Õ ( Capital letter O with tilde or O-tilde ) | O |  |
| ASCII code 230 = µ ( Lowercase letter Mu ; micro sign or micron ) | **U** |  |
| ASCII code 231 = þ ( Lowercase letter Thorn ) | **P** |  |
| ASCII code 232 = Þ ( Capital letter Thorn ) | **P** |  |
| ASCII code 233 = Ú ( Capital letter U with acute accent or U-acute ) | **U** |  |
| ASCII code 234 = Û ( Letter U with circumflex accent or U-circumflex ) | U |  |
| ASCII code 235 = Ù ( Capital letter U with grave accent ) | U |  |
| ASCII code 236 = ý ( Lowercase letter y with acute accent ) | Y |  |
| ASCII code 237 = Ý ( Capital letter Y with acute accent ) | **Y** |  |
| ASCII code 238 = ¯ ( Macron symbol ) | **0** |  |
| ASCII code 239 = ´ ( Acute accent ) | **0** |  |
| ASCII code 240 = ≡ ( Congruence relation symbol ) | **0** |  |
| ASCII code 241 = ± ( Plus-minus sign ) | **0** |  |
| ASCII code 242 = ‗ ( underline or underscore ) | **0** |  |
| ASCII code 243 = ¾ ( three quarters, three-fourths ) | **0** |  |
| ASCII code 244 = ¶ ( Paragraph sign or pilcrow ; end paragraph mark ) | **0** |  |
| ASCII code 245 = § ( Section sign ) | **0** |  |
| ASCII code 246 = ÷ ( The division sign ; Obelus ) | **0** |  |
| ASCII code 247 = ¸ ( cedilla ) | **0** |  |
| ASCII code 248 = ° ( Degree symbol ) | **0** |  |
| ASCII code 249 = ¨ ( Diaresis ) | **0** |  |
| ASCII code 250 = · ( Interpunct or space dot ) | **0** |  |
| ASCII code 251 = ¹ ( Superscript one, exponent 1, first power ) | **0** |  |
| ASCII code 252 = ³ ( Superscript three, exponent 3, cube, third power ) | **0** |  |
| ASCII code 253 = ² ( Superscript two, exponent 2, square, second power ) | **0** |  |
| ASCII code 254 = ■ ( black square ) | **0** |  |
| ASCII code 255 = nbsp ( Non-breaking space or no-break space ) | **0** |  |

Reference

http://www.theasciicode.com.ar/ascii-control-characters/end-of-text-hearts-card-suit-ascii-code-3.html