**CA478 Character Sets MP for BAHv2 for MX messages?**

As you know, some intermediaries have been confronted lately to the use of non-basic latin characters in the ISO 20022 SID or GM messages.

In most cases it seems to break STP since current back office or interface systems do not support more than the ASCII or EBCDIC (Basic latin) characters & the processing must be done manually.

One of the big global custodian asked me if the “CharacterSet” element of the BAH could be used to indicate the character set(s) that would be used in the messages when non-basic latin characters are used. This would not really solved the problem short term but at least it could provide a way to filter out messages with special characters.

**Could the SMPG do a market practice on the usage of that field to specify the six ISO character sets values that could be populated in the context of SRDII (see Table below)?**

It is true that it is not at all obvious to find these character set values since they are not part of the schemas and practically this field corresponds to an unlimited narrative field today. This field could eventually be standardised by ISO in the future when discussing BAH v3 by defining a finite code list?

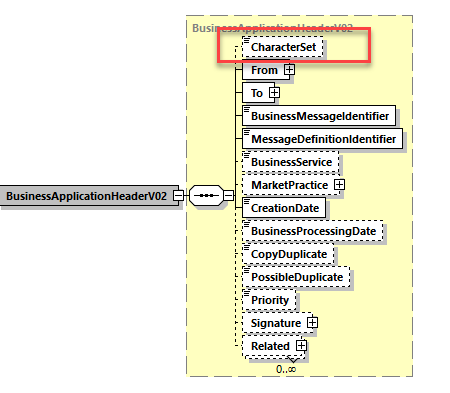


Figure 1: Position of the Character Set field in the BAHv2 - head.001.001.02)

As defined in the ISO 20022 repository, the character set field values must be UNICODE “Code Block Names” and the values can be found here: <http://unicode.org/Public/UNIDATA/Blocks.txt>

|  |  |  |
| --- | --- | --- |
| ISO Character sets  (Beyond Basic Latin) | **UnicodeChartsCode** – **Code block name** | Regions in which the character set is used |
|  | *Code list Values for the BAHv2* ***CharacterSet*** *element* |  |
|  | *(casing, whitespace, hyphens, underbars are ignored)* |  |
| ISO 8859-1 Latin 1 | **Latin-1 Supplement** | Western Europe |
| ISO 8859-2 Latin 2 | **Latin Extended-A** | East European (eg. Slovac) |
| ISO 8859-3 Latin 3 | **Latin Extended-A** | South European (eg. Maltese) |
| ISO 8859-4 Latin 4 | **Latin Extended-A** | North European (eg. Latvian) |
| ISO 8859-5 Cyrillic | **Cyrillic** | Bulgary |
| ISO 8859-7 Modern Greek | **Greek and Coptic** | Greece |

ISO Character set names corresponding to each EEA Language:

|  |  |  |
| --- | --- | --- |
| **Language** | **ISO 639-1 code** | **Character set** |
| Bulgarian | BG | ISO 8859-5 Cyrillic |
| Czech | CS | ISO 8859-2 Latin-2 |
| Danish | DA | ISO 8859-1 Latin-1 |
| German | DE | ISO 8859-1 Latin-1 |
| German | DE | ISO 8859-1 Latin-1 |
| Greek | EL | ISO 8859-7 Modern Greek |
| English | EN | ISO 8859-1 Latin-1 |
| Spanish | ES | ISO 8859-1 Latin-1 |
| Estonian | ET | ISO 8859-4 Latin-4 |
| Finnish | FI | ISO 8859-1 Latin-1 |
| French | FR | ISO 8859-1 Latin-1 |
| Irish | GA | ISO 8859-1 Latin-1 |
| Croatian | HR | ISO 8859-2 Latin-2 |
| Hungarian | HU | ISO 8859-2 Latin-2 |
| Icelandic | IS | ISO 8859-1 Latin-1 |
| Italian | IT | ISO 8859-1 Latin-1 |
| Lithuanian | LT | ISO 8859-4 Latin-4 |
| Latvian | LV | ISO 8859-4 Latin-4 |
| Maltese | MT | ISO 8859-3 Latin-3 |
| Dutch | NL | ISO 8859-1 Latin-1 |
| Norwegian | NO | ISO 8859-1 Latin-1 |
| Polish | PL | ISO 8859-2 Latin-2 |
| Portuguese | PT | ISO 8859-1 Latin-1 |
| Romanian | RO | ISO 8859-2 Latin-2 |
| Slovak | SK | ISO 8859-2 Latin-2 |
| Slovene | SL | ISO 8859-2 Latin-2 |
| Swedish | SV | ISO 8859-1 Latin-1 |