European
Payments Council

[**X**] Public – [] Internal Use – [ ] Confidential – [ ] Strictest Confidence
Distribution: N/A

SEPA REQUIREMENTS FOR AN EXTENDED CHARACTER SET (UNICODE SUBSET) BEST PRACTICES

Abstract	This document contains a set of recommended best practices to be used in dealing with local language and special characters used in SEPA countries
Document Reference	EPC217-08
Issue	Version 1.1
Date of Issue	18 December 2009
Reason for Issue	Approval by the EPC Plenary 16 December 2009
Reviewed by	The Standards Support Group
Produced by	The Standards Support Group
Authorised by	The Standards Support Group
Circulation	Publicly available

TABLE OF CONTENTS

0. **Management Summary**

Background: Within the SEPA schemes, unless otherwise agreed, only a limited set of Latin characters may be exchanged. This restriction has been agreed in order to overcome the complexity that would result if all European language based characters were to be exchanged between all adhering banks. However banks by prior agreement may exchange additional characters such as through bilateral/multilateral agreements or via an agreed AOS (Additional Optional Service).

Emerging implications: Although restricting SEPA to a limited set of Latin characters has avoided complexity and minimised the impact on bank processing systems it has other less positive consequences. Specifically these are that end users may not be able to convey, via the basic SEPA schemes, local language characters end-to-end. Broadly the two key business drivers are:

- that the adoption of SEPA instruments in a domestic setting places greater emphasis on the requirement to cater for local language characters;

- and that large corporate users require certainty that payment related data will be conveyed end-to-end irrespective of the European characters it may contain.

It also needs to be pointed out that UNICODE UTF-8, the character set underlying the ISO 20022 XML message standards, supports all language characters, although it is not expected that the full UNICODE UTF-8 will be implemented by adhering banks. Moreover even before the introduction of ISO 20022, users commonly 'filter' incoming messages to ensure that only recognised characters are in use and or convert them for use in various legacy systems.

The Recommended Solution: The solution provides a common conversion tool (SEPA conversion table) to optionally extend the Latin character set to cater for all European local languages. The tool may be used to protect legacy systems, and it is accompanied with a set of best practices covering the intended usage of the conversion table within SEPA.

The best practice allows the original sender of a message to use, optionally and only by prior agreement, local language and special characters in text fields (e.g., name, address and remittance information). The sending bank then has the responsibility to ensure that the subsequent receivers of local and special characters can process and forward these fields unchanged.

1. Document Information

1.1. Change History

This is the first release of this document. As necessary, further versions may be issued to address developments relating to changes in the ISO standards or market practice.

Issue number	Dated	Reason for revision
V1.0 Draft	May 2009	For national banking consultation
V1.0 Draft	November 2009	For approval for publication
V1.1	December 2012	Section 6.1: Include a hyperlink to the conversion table document Update of the EPC address

2. Introduction

This document describes

- the underlying issues related to the use of local language and special characters (such as those used in internet URLs and financial texts) in the context of the SEPA payment schemes which are based on the ISO 20022 XML messages

- the solution which consists of a recommended conversion table and best practices on its use.

3. Background on the need for action and related issues

The current EPC Implementation Guidelines define a basic Latin character set that consists of the following characters:

a b c d e f g h i j k l m n o p q r s t u v w x y z

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

0 1 2 3 4 5 6 7 8 9

/ - ? : ( ) . , ' +

Space

The limited character set reflects the need to ensure information can be, for example, printed on statements.

The EPC Implementation Guidelines allow the use of local characters under pre-arranged agreements, as for example, in a Greek language community.

The ISO 20022 XML message standards underlying the SEPA schemes support all UNICODE characters and would thus cater for all SEPA language requirements and any other needs, such as @, #, *, and € which are used in modern business communications.

The current practice is that in lieu of rejecting a payment, some receiving banks replace unrecognised characters with blanks or other characters or simply delete them. This indicates that there is no common practice in place and introduces inefficiencies and uncertainty in processing.

Furthermore, the limited character set is used to facilitate the transfer of information used in different standards to legacy messages such as the SWIFT-reporting statements.

Some corporates[1] report that they have databases which have names, addresses and other text information in a local character set[2]. They have therefore requested to be able to use these without conversion for SEPA payments.

The proposed solution facilitates the continued use of local language and special characters, by defining a broader character set and by providing a conversion table and best practices. It does not however mandate this solution upon the entire community as the impact and cost of doing so would be very significant.

---

[1] A member of the group is a representative of EACT and the need to address the character set problem has been supported by the EPC Customer Stakeholders Forum held 18 December 2008

[2] These data bases also include names and addresses in non-European language characters, but the SEPA characters set would not need to cater for these

4. Situation today on the Character Set in SEPA

Local language characters in one community are not always supported in another community. Technical constraints[3] have so far made it impossible to come up with a single solution to support 'domestic' and 'cross-border' transactions.

In removing the distinction between 'domestic' and 'cross-border' transactions, the expectation is to support the full set of characters used in the language of each SEPA country as made possible in the use of ISO 20022 XML message standards. To facilitate this, the different existing code pages[4] need to be aligned and legacy systems eventually upgraded. In view of the large number of code pages embedded in software and hardware, the migration will have a pervasive impact on both bank and end-user systems. The filters can be revised based on the proposed conversion table where there is a business need to map incoming and outgoing data in order to protect the legacy systems. Until such time community wide migration to UNICODE UTF-8 is not realistic.

The need to deal with local language characters may occur at different stages in the payments chain and for a variety of reasons. For example, in a credit transfer:

- A receiving bank may need to store, display or print the name and address information, potentially requiring conversion to the Latin character set.

- The receiving bank may be able to process, store or print the name or address in the local character set used in the original message. This can only be achieved if all intermediary parties are able to forward the original information as received.

Another request to expand the EPC basic Latin character set relates to **special characters** in use to identify e-mail addresses, URLs and legal names of parties, eg, '@', '!', '#', '&' that would be used in names, remittance and other information. Special attention will be required for the '€' and '@' characters which have been implemented in different ways and for which a more sophisticated solution may be required to avoid disrupting legacy systems.

5. Recommended best practice

The solution provides a **common SEPA conversion table** which enables conversion to a restricted Latin character set and can be used, for example, for printing in the statements or for additional processing in legacy environments.

The solution applies to selected data elements as identified in section 6.3

It is recommended that banks use the common conversion table when necessary to minimise the risk of rejection.

In using the common conversion table, the following best practices apply:

- **first principle** – normal banking practice applies, where the relationship between a bank and its customer is in the competitive space, based on a contractual agreement. Therefore banks are free to decide whether they will support local language and special characters. Nevertheless it should be noted that it is the responsibility of the originating customer to ensure in conjunction

---

[3] For example, mainframes and processing rules have required the use of the basic Latin character subset in cross-border transactions.

[4] A code page is a means of providing support for character sets and keyboard layouts for different countries or regions. A code page is a table that relates the binary character codes used by a program to keys on the keyboard or to characters on the display.

with originating bank, that the receiving customer and the receiving bank can handle the information.

- **second principle** – each subsequent receiver, acting as intermediary, is responsible for forwarding information as received from the previous sender (i.e., forward text unchanged, regardless of any processing or conversion done), unless otherwise agreed.

- **third principle** – for processing the received information:

  o when the receiver has the technical capability, he should process the information as received without alteration/conversion

  o if the receiver does not have the capability, he should apply the SEPA conversion table to process the received information.

  o the receiver can only reject a message if it contains characters that are not contained in the conversion table which applies to SEPA country requirements. The sender remains responsible for sending messages containing supported characters as shown in the conversion table so as to minimise rejects.

- **fourth principle** – each sender should accept that returned messages may include information resulting from a conversion according to the SEPA conversion table. This means that the sender of the original transaction may need to link original information and returned converted information. The returning party may not be held responsible for differences between the original and converted data or any truncation in case of one-to-many character conversion.

In agreement with their banks, customers will be free to use local language characters as specified in the conversion table, as currently done in domestic payment environments thus supporting the creation of a borderless SEPA payments environment that supports all local language characters used in the different SEPA countries.

6. The SEPA conversion Table

6.1. How it is to be used

The conversion table provides a common way for banks to convert characters when and if needed in handling SEPA transactions. (Click on following link to download the spreadsheet containing the conversion table EPC217-08_SEPA_Conversion Table V_10.xls) This table is based on international rules for romanisation, in particular ISO 9 and ISO 843 and input of the Greek community.

The conversion table is defined as the set of yellow, green and red entries. The aim is to include all recognised European language requirements and text requirements as used in financial texts while ensuring easy implementation by having as few exceptions as possible, where:

- Yellow entries are those specified in the EPC Implementation Guidelines as the Basic Latin Character Set, (BLCS)

- Green entries are those used in European languages in addition to the BLCS, for example,

  o Greek, Cyrillic and other European characters

  o The cent, pound, plus-minus, euro sign, (but not the generic currency symbol) were accepted as part of the SEPA Character Set as they have uses in a financial context

- The inverted '!' and '?' for example were accepted, more as a matter of practical use
- The tilde was accepted as it is used in URLs – see also section 6.2 that suggests e-mail addresses are likely to be usable once converted
- Red entries are characters that are outside the extended character set, eg, Chinese and Arabic characters. The presence of these characters can be the basis for a reject or be replaced by a full stop (U+002E).

The table provides the following information per character

- A 'picture' of the character or sign (Col A), its denotation or name (Col B) and the Code point in Unicode.[5] (Col C)

- A basic conversion table (Col D and Col E), where
  - 'none' denotes that the character is already part of the Basic Latin Character Set
  - 'NA' means that conversion is not possible
  - '.', a full stop (U+002E) denotes that there is no restricted character set equivalent and hence a full stop is to be used
  - another character in the restricted Basic Latin Character Set is provided as a substitute, such as many of the basic Greek and Cyrillic characters, or as another example, right square bracket is replaced by a right parenthesis

- An extended conversion table (Col F and Col G), including the additional characters (green entries) to be added to the BLCS

- Countries known to use the local language character (Col I)

- If the sign is a special XML character (Col H) – see section 6.2 below

- To minimise truncation, the table focuses on a one character to one character conversion. Whenever a one to multi-character conversion is unavoidable, a warning note is provided in the table (Col H), In such cases where there are length limitations, it is recommended that the sender uses the conversion table to count the number of characters that would result from conversion at the receiving end. In any event, the converting party cannot be held responsible for any resulting truncation resulting from SEPA requirements.

- To ensure that the responsibilities of the sending and receiving parties are held, the SEPA scheme rulebooks should refer to this paper. This would minimise the risk for mis-interpretations on the use of the extended character set within SEPA.

---

[5] Note that because of the UTF-8 encoding, the Code point value is not the same as the binary presentation of the sign in the XML message.

6.2. Handling of Special Characters

In expanding the character set for special characters, the following rule applies for the sender:

Rule: Senders are to ensure that the following five XML special characters may only be used with escape characters, as follows:

Special XML Character	Representation
& (ampersand)	&
< (left brace)	<
> (right brace)	>
" (double quotation)	"
' (single quotation)	'

For the sender, best practice solutions for urgently needed special characters, which are not in the restricted character set, are as follows:

Special Character	Representation
€	E (capital E)
@	(at)
&	+ (plus)
_	- (hyphen)

The final solution is that the characters are fully supported as part of the extended Character Set.

6.3. Rules for selected ISO 20022 XML data elements

The general rule is that data elements that are processed by computer systems must be limited to the Latin character set. It should be noted that reference data element which often need to be stored or processed by the receiving bank or clearing systems usually follow the international practice, namely the use of the restricted Latin character set. Examples of this are data elements containing amounts, codes and dates and references. Data elements that need to be read by humans such as name and address information and free text, may contain local language characters.

Text data elements in which special local characters can be expected to appear include many important **data elements**, such as:

- Name and Address
- Remittance and other Information data elements
- Instruction for Creditor Agents and other parties

Special rules as noted above apply to all reference or identification data elements that may be used by the receiver in processing the message:

- Instruction Identification
- End-to-End Identification
- Transaction Identification

- Message Identification

- Payment Information Identification

- Creditor and Debtor Identification

- Ultimate Debtor/Creditor Identification

- Remittance Identification

- Proprietary codes – if used in processing in the payment/banking chain

Rule: Senders must ensure that the content of Identifiers/reference data elements

- Is restricted to the restricted basic Latin character set

- Must not start or end with a '/'

- Must not contain two consecutive '/'s anywhere in the data element