

- 普段利用する言語が異なるもの同士がやり取りする以上、文字としては国際的な共通言語である英語（英アルファベット＋数字＋記号）を利用するものとする。
- 各言語で特有の文字（例えば漢字）を利用するのであれば、Unicodeを使うのが自然と考える。
- その観点で、サポートする文字コードとして、ASCII、UTF-8、UTF-16、EBCDICで記述されたデータを読み書きできることが必須。
- やり取りする前／した後のデータの文字コードは、限定しない（各組織それぞれのニーズによる、相互変換が必要）。

文字コード	概要
ASCII	<ul style="list-style-type: none">•ANSI（アメリカ規格協会）の前身であるASAによって1963年に定められた、英語のアルファベットや数字などを中心とする文字コード体系。•7ビットの情報で表現されており、128種類の文字が含まれている。このうち33の文字は、制御文字。•ASCIIの規格は米国で策定され、後にISOによって「ISO 646」として国際標準化された。•日本では、「ISO 646」に基づく日本工業規格として「JIS X 0201」が策定された。 <p>ASCIIを国際化してアメリカ以外でも使えるようにしたのが、ISO646。具体的にはASCIIの図形文字94文字中、アメリカでのみ使われる12文字を各国で自由に決めていいものと規定。（共通なのは82文字） ISO646の日本語版が、「JIS X 0201」に相当。 それぞれの国、地域でASCIIというとISO646を指していることが殆どと考えられるので、考慮が必要（双方で表示文字が変わる）</p>
UTF-8	<ul style="list-style-type: none">•Unicodeを8ビットの可変長マルチバイトでエンコーディングする方式。•英数は1バイトで表現し、日本語は3バイトで表現。•UTF-8方式を用いて文字列を変換すると、Unicodeの最初の128文字を変換した結果がASCIIコードと全く同じになるため、旧来の処理システムとも親和性が高く保つことができる。
UTF-16(BE/LE)	<ul style="list-style-type: none">•Unicodeを16ビットを単位とした可変長マルチバイトでエンコーディングする方式。•Unicodeの2バイトで定義されているコード（U+0000～U+D7FFF、U+E000～U+FFFF）はそのまま、サロゲートペアで定義されているコードは4バイトにエンコード。•ビッグエンディアン用のUTF-16BE、リトルエンディアン用のUTF-16LEがある。
EBCDIC	<ul style="list-style-type: none">•IBM社によって開発された、主にIBM系のメインフレームで採用される文字コード。•256個の文字、数字と記号の表示が可能。•1つの文字は8ビットのバイナリ数字として表現。 <p>•EBCDICには様々な拡張がある。 •日本語用EBCDICとしては、一般的にEBCDIKと呼ばれる、空き領域にカタカナを追加したものがある。</p>

参考：EDIFACTで利用可能な文字セット

利用されるデータ表現形式としては、EDIFACT, X12, XMLでほとんどを占めると思うので、それぞれの文字コードの取り扱いを簡単に整理した。

EDIFACTでは、メッセージヘッダーにあたるUNBセグメントに、メッセージで利用する文字セットをキーワード指定する。

Ex. UNB+**UNOC**:3+5012345678901:14+4598765432198:14+000316:1402+INV73529++INVOIC'

指定可能な文字セットは以下の表のとおり

キーワード	エンコーディング
UNOA	ISO 646 (a...z の英子文字を除く) . , - () / = (space)
UNOB	ISO 646 All of UNOA ' + : ? ! " % & * ; < >
UNOC	ISO 8859-1 (Part 1: Latin alphabet No. 1)
UNOD	ISO 8859-2 (Part 2: Latin alphabet No. 2)
UNOE	ISO 8859-5 (Part 5: Latin/Cyrillic alphabet)
UNOF	ISO 8859-7 (Part 7: Latin/Greek alphabet)
UNOG	ISO 8859-3 (Part 3: Latin alphabet)
UNOH	ISO 8859-4 (Part 4: Latin alphabet)
UNOI	ISO 8859-6 (Part 6: Latin/Arabic alphabet)
UNOJ	ISO 8859-8 (Part 8: Latin/Hebrew alphabet)
UNOK	ISO 8859-9 (Part 9: Latin alphabet)
UNOX	ISO 2022-JP (JISコード)
UNOY	ISO 10646 (Unicode)

ISO 8859 はヨーロッパの文字セット。
例えば、8859-1 は、フランス語、ドイツ語といった西ヨーロッパ語が書ける。

参考：X.12で利用可能な文字セット

- X.12では、以下の文字セットが利用可能と規定されている
- エンコーディングとしては、「ASCII」と「UTF-8」が推奨されているよう。

基本文字セット	A...Z	0...9	!	"	&	'	()	*	+
	,	-	.	/	:	;	?	=	スペース	
拡張文字セット	a...z	%	~	@	[]	_	{	}	¥
		<	>	#	\$					

参考：XMLで利用可能な文字セット

- XML文書の先頭に記述するXML宣言中に、XML文書がどのような文字コードで記述されているかを明示する。

Ex. `<?xml version="1.0" encoding="UTF-8" standalone="yes"?>`

- Encoding属性に記述できる値（名称）は、IANAに登録されたキャラクタセット名を使用する。
 - IANA (Internet Assigned Numbers Authority)は、インターネットにおいて、名前や番号の登録を必要とする情報に関して、登録業務を行う組織
 - キャラクタセット一覧： <http://www.iana.org/assignments/character-sets/character-sets.xhtml>