



SLOVENSKI STANDARD SIST-TP CR 13928:2003

01-oktober-2003

Informacijska tehnologija – Vodilo za uporabo standardov za znakovne nabore v Evropi

Information Technology - Guide to the use of character set standards in Europe

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Ta slovenski standard je istoveten z: CR 13928:2000

<https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003>

ICS:

35.040

Nabori znakov in kodiranje
informacij

Character sets and
information coding

SIST-TP CR 13928:2003

en

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST-TP CR 13928:2003

<https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003>

CEN REPORT
RAPPORT CEN
CEN BERICHT

CR 13928

May 2000

ICS

English version

Information Technology - Guide to the use of character set
standards in Europe

This CEN Report was approved by CEN on 12 April 2000. It has been drawn up by the Technical Committee CEN/TC 304.

CEN members are the national standards bodies of Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[SIST-TP CR 13928:2003](https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003)

<https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003>



EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

Central Secretariat: rue de Stassart, 36 B-1050 Brussels

FOREWORD

This report was produced by a CEN/TC 304 Project Team, set up in June, 1998, as one of several to carry out the funded work program of TC 304 (documented in CEN/TC 304 N 666 R2). A first draft was discussed at the TC meeting in Brussels in November, 1998. A revised draft was circulated for comments within the TC and thereafter discussed at the TC plenary meeting in April, 1999. This revised version is based upon comments received during and after that meeting and is circulated for written ballot within the TC. This report was approved by the TC. It is sent now to the CEN BT for approval.

iTeh STANDARD PREVIEW (standards.iteh.ai)

[SIST-TP CR 13928:2003](#)

<https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003>

TABLE OF CONTENTS

FOREWORD	2
1 INTRODUCTION	5
2 EXECUTIVE SUMMARY	5
3 SCOPE AND FIELD OF APPLICATION	6
4 DEFINITIONS	6
5 CHARACTERS AND THEIR CODING	7
5.1 Characters, glyphs and languages	7
5.2 Coding	7
5.3 Control functions and control characters	8
6 THE CHARACTER HANDLING MODEL	8
6.1 The input function	9
6.2 The processing function	9
6.3 The interchange function	10
6.4 The output function	10
6.5 Cultural issues	11
7 OFFICIAL STANDARDS, MANUFACTURER STANDARDS, AND RELATED STANDARDS	11
7.1 Telecommunication standards	11
7.2 Manufacturer standards	12
7.3 Related Standards	12
8 INTERNATIONAL CHARACTER SETS	12
8.1 Framework standards for 7- and 8-bit environments	13
8.2 7- and 8-bit character set standards	14
8.3 The universal character set (UCS) standard	15
8.4 Control functions	16

iTech STANDARD PREVIEW

(standards.itteh.ai)

[SIST-TP CR 13928:2003](https://standards.itteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-153cf3a09820/sist-tp-cr-13928-2003)<https://standards.itteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-153cf3a09820/sist-tp-cr-13928-2003>

8.5	Character description standards	17
9	EUROPEAN CHARACTER SETS	17
9.1	8-bit character sets	17
9.2	16-bit character sets – The multilingual European subsets	18
9.3	The EURO SIGN	19
10	PROCUREMENT ISSUES	19
10.1	Repertoires and code structures	19
10.2	Transformation and fall-back	20
10.3	Code structure interoperability	22
11	PROCUREMENT CLAUSES	22
11.1	Structure	23
11.2	Input character repertoire	23
11.3	Output character repertoire	23
11.4	Processing character repertoire	24
11.5	Interchange character repertoire	25
11.6	Additional requirements when using the 8-bit code structure for interchange	27
11.7	Additional requirements when using the multi-byte UCS code structure for interchange	27
12	CEN AND CEN/TC 304	28
13	REFERENCES	28
13.1	ISO/IEC standards	28
13.2	European standards	30
13.3	ITU-T Standards	30
13.4	ETSI standards	30
13.5	Unicode	30
13.6	Other non-standard references	31
Annex A	8-BIT CHARACTER SETS	32
Annex B	THE UNIVERSAL CHARACTER SET (UCS)	71

iTech STANDARD PREVIEW
(standards.iteh.ai)

[SIST-TP CR 13928:2003](https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003)

<https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003>

1 Introduction

There exist today a large number of standards and related specifications concerning character repertoires and their coding in the form of official as well as manufacturer standards and intended for a wide range of applications and uses. Furthermore, there are character set standards for data communication and there are standards developed specifically for telecommunications applications. The situation can be very confusing to the non-expert user and to people involved in procurement.

The user of IT systems normally does not have to be concerned with these types of standards. However, there may be situations where the user has to be able to express working needs for certain character repertoires. It may also happen that the user, when involved in work together with other parties using other systems, needs to be able to interpret other people's specifications given in the form of reference to standards.

The procurer of IT systems should be able to specify requirements in the form of reference to established standards.

A particular purpose of the report is to give guidance for public procurement in Europe. Since there is an EC directive and a council decision for the use of official European standards in procurement above certain monetary amounts, the report concentrates on such standards. There may be future editions, in which case more attention will be given other types of standard. (See also section 7.)

The main purpose of this report is to give guidance to users and procurers by explaining the purposes and relationships of the official standards in the domain of data communication. Explicit guidance is given in paragraphs marked with ▶.

The text is presented on two levels. The first level, contained in the body of the report, provides a general coverage of character repertoires, coding and uses. The second level, contained in the two annexes, provides much more detailed, tutorial information. The reader who finds the level of technical detail too deep

may be better served by the "Manual: Standards for the electronic interchange of personal data: Part 5 – Character sets" (see References).

Further information on character sets and their standardization can be found in the document "Language automation world-wide: The development of character set standards" and on the Letter Database web site (see References).

2 Executive summary

The main body of this report is aimed primarily at the non-technical person who needs to become familiar with the use of character set standards in Europe for various purposes in an IT environment. This audience will include managers/decision makers and their advisors; administrators (for procurement purposes); technicians (for programming and system development purposes); standardisers; perhaps also journalists.

The concepts of characters and their coding is introduced in section 5, and a conceptual model on the use of coded character sets is provided in section 6. The guide concentrates on official character set standards. However, there is a range of other standards for character sets that are not official, and there are also specifications concerning associated topics such as rules for ordering character strings. Section 7 goes on to place the official standards in the wider context of these other standards. Sections 8 and 9 describe a range of official character set standards with an international and a European scope respectively. Section 10 introduces a number of procurement issues, and section 11 provides sample text that may be used as the basis for inclusion in (public) procurement specifications for IT systems and software.

In addition, the guide has two annexes which contain a much more technical description of official character set standards.

The activities of CEN/TC304, the committee responsible for the promulgation of character set and related specifications in Europe, are described in section 12, and finally pointers for further reading and research are given in section 13.

3 Scope and field of application

The technical scope of this guide is primarily limited to official character set standards promulgated by ISO/IEC and CEN, as opposed to official telecommunications standards and manufacturer standards. However, an overview of all types of standards is given in section 7. The guide furthermore concentrates on European issues; thus character set standards for non-European languages are not covered.

The guide is mainly intended as an introduction for people who need to familiarise themselves with the concept of character sets and their coding; e.g. managers/decision makers and their advisors; administrators (for procurement purposes); technicians (for programming and system development purposes); standardisers; perhaps also journalists. Particular emphasis is placed on its use by procurers.

4 Definitions

The following terms are used in the body of this report and the official definitions are given here where they exist. They are taken from the standards ISO/IEC 9541:1991 and ISO/IEC 10646-1:2000, except when denoted by an * where a definition is created for the purposes of this guide.

4.1 character: A member of a set of elements used for the organisation, control, or representation of data.

4.2 (character) repertoire: A specified set of characters that are represented in a coded character set.

4.3 *code table: A tabular representation of a coded character set, showing also the coded representations.

4.4 *code page: Synonym for code table, used in the IBM PC environment.

4.5 *code space: The numeric domain occupied by all bit combinations used for the coding of a coded character set.

4.6 coded character set: A set of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their coded representation.

4.7 combining character: A member of an identified subset of a coded character set, intended for combination with the preceding or following graphic character, or with a sequence of combining characters preceded or followed by a non-combining character

4.8 control character: A control function the coded representation of which consists of a single bit combination.

4.9 control function: An action that affects the recording, processing, transmission, or interpretation of data, and that has a coded representation containing one or more bit combinations.

***Note –** A bit combination in this context is a 7- or (more commonly) 8-bit byte but in the case of UCS-2 and UCS-4 the bit combinations are 16 and 32 bits respectively.

4.10 *diacritic, diacritic mark: A mark intended for the association with a letter (e.g. acute accent).

4.11 *fall-back: A non-reversible transformation consisting of the substitution of an input character which cannot be represented on the output device by one or more characters which can.

4.12 glyph: A recognisable abstract graphic symbol which is independent of any specific design.

4.13 graphic character: A character, other than a control function, that has a visual representation normally hand-written, printed, or displayed.

4.14 transliteration: The process which consists of representing the characters of an alphabetical or syllable writing system by the characters of a conversion alphabet.

Note – In principle, a transliteration should be a one-to-one conversion.

5 Characters and their coding

5.1 Characters, glyphs and languages

For the presentation of written text we use letters, digits and punctuation marks. Often we also use special symbols such as currency signs. All of these are called graphic characters, and the collection of characters for a specific purpose, such as the presentation of text in a specific language, is called in the standardisation context a *character repertoire* (or *character set*). The most common type of repertoire is of course the alphabet of a language, complemented by the ten digits and a set of special characters.

A graphic character is represented in printed form or on a display surface; hence it must have an agreed shape. Of course, a character may be represented by many variations of its basic shape (or shapes, as with g and g) depending on the font in use (e.g. Times Roman or Arial). No matter how many such variations may be used to represent a character, the basic shape is usually recognisable to the human eye. This inherent shape of a character, which is independent of font, is known as a *glyph*. However, it should be recognised that this concept is less straightforward than it first might appear. Thus one and the same glyph may represent, in different contexts, different characters (e.g. the Latin character B is not the same as the Cyrillic character Б).

Although the glyph concept is important for the definition of character repertoires, it is not central to the theme of this guide. The reader who wishes to obtain more information about glyphs is referred to ISO/IEC TR 15285, *An operational model for characters and glyphs* (see References).

Almost every language has its own character repertoire. However, the fact that many European languages have a large number of characters in common naturally facilitates the work on defining character repertoires for Europe. In CEN/TC 304 there is a separate activity on providing a catalogue of the alphabets of indigenous languages.

5.2 Coding

In IT systems a character is represented by a 7-, an 8-bit, a 16-bit or a 32-bit combination, usually expressed as a numeric code. A character repertoire with its corresponding set of codes is called a *coded character set* or just *codeset*. Such a set is often represented graphically in the form of a *code table* (Figure 1), which also illustrates the principles of the distribution of the codes, the *code structure*. Furthermore, the totality of the bit combinations used for a coded character set is called its *code space*.

	0	1	2	3	4	5	6	7
0			SP	0	@	P	`	p
1			!	1	A	Q	a	q
2			"	2	B	R	b	r
3			#	3	C	S	c	s

Figure 1 – Code table. The first four rows (out of 16) of a 7-bit code table. The row number translated into binary form gives the four least significant bits of the bit combination; the column number gives the three most significant bits.

Coded character sets are used for different purposes in computer systems, and the code structures may therefore vary. For instance, a coded character set used for interchange purposes often needs codes to be reserved for control characters, so that these may be included in the interchange data stream. However, a coded character set used for processing purposes may not need such reserved areas, which instead are often only used to represent more graphic characters.

5.2.1 Proliferation of codes; standardisation

Early IT systems had severe size limitations. Therefore, the character codes had to be kept small. The earliest codes occupied 5 and 6 bits; later 7 and 8 bits have been used. These provide a coding capacity for 32, 64, 128 and 256 characters respectively. However, even with an 8-bit representation it is not possible to support all European languages in a single coded

character set. Thus coded characters sets proliferated. As long as an application (and the character set it used) was restricted in use to a single country or geographical region, this proliferation did not create problems, since a character set could be chosen to support the limited number of languages for that region. However, due to the requirements of international trade and the increase in travel, the limitation in the number of characters in one coded set has caused great problems of application interoperability.

In order to avoid a very large number of private character set specifications, many with overlapping scope and leading to interoperability problems, standardisation was needed. It was carried out both by the official standardisation organisations and by the manufacturers, most notably by IBM, Apple and Microsoft.

Modern IT systems no longer have the earlier restrictions in size, and a solution using 16 and 32 bit encoding is now available which uses a code space sufficient to accommodate the characters of every language in the world in one and the same coded character set. However, since the old solutions seem likely to continue to exist until perhaps 2025, the old problem may remain acute for some time. There will be the added complication of using the old and the new systems together as well as how to migrate, in an orderly fashion, to the new system.

5.3 Control functions and control characters

For IT processing purposes, it is necessary to indicate within a data stream where some action is required, e.g. a carriage return or new line. Such actions are performed through *control functions*, which do not have graphical representations. Over 160 control functions have been standardised. Some of them, such as the carriage return, are represented by a single *control character*, which has a coded representation and can therefore be included in a code table. Others are represented by a sequence of characters with a special introducing control character at the beginning of the sequence.

6 The character handling model

Figure 1 below illustrates the character handling model. It represents a simplified IT scenario which consists of two computer systems connected by a communications link. The purpose is to show the different aspects of the handling of characters by users and computer systems and thus introduce basic concepts that will be used in the following sections of this guide. It is also intended to help differentiate between the roles of the user(s) and the procurer in the context of this guide.

Note – In parallel with the development of this guide, a separate CEN Technical Report was produced entitled “Character Repertoire and Coding Transformations – General model for graphic character transformations”. The model described in this guide is entirely consistent with the more general model.

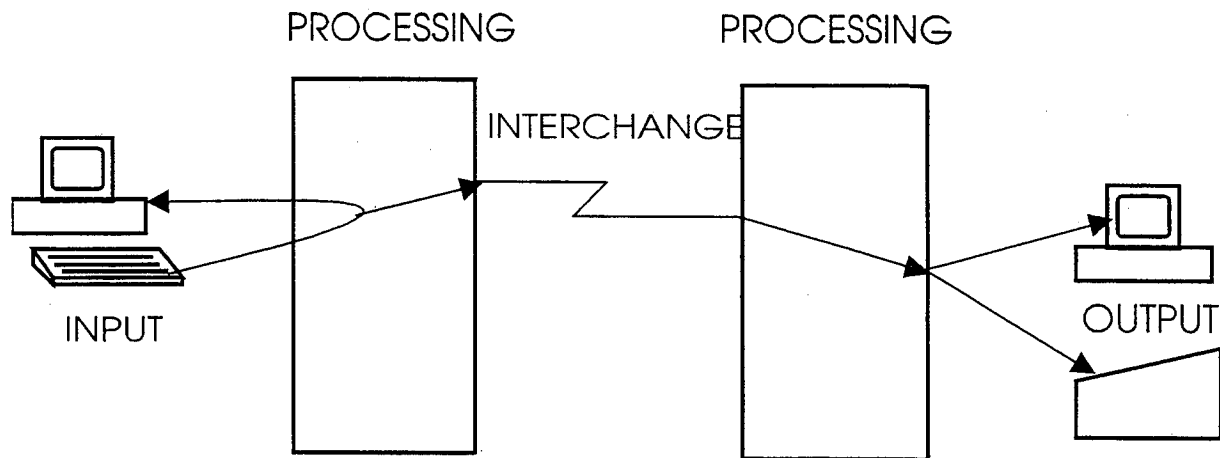


Figure 2 – Character Handling Model

6.1.1 Keyboards

6.1 The input function

The input function provides for the entering of data into a computer system. Figure 2 uses a keyboard for input, but any device capable of entering character data may be used.

For more complicated input, for example for input of ISO 10646 characters, an input method is needed, that is a program is assisting the choice of the correct character after one or more characters have been entered from the keyboard. This may be in the form of a list of characters to choose from, displayed on the screen, or a specific sequence of characters may be entered after activation of a special key, like a "compose" key, and then some name or mnemonic characters for the key.

- For the user, the main issue is whether or not there is available in the computer system a character repertoire for input which is sufficient for the requirements. For the procurer, the main issue is to produce a procurement specification which satisfies the input needs of all intended users of the product.

Note that the representation of the input text on the monitor screen is a result of both the processing function, e.g. a word processor, and an output function (to the screen).

The main keyboard standard is ISO/IEC 9995, *Keyboard layouts for text and office systems*. Keyboard standards are related to character set standards but are not central to the National keyboard standards have, in general, been promulgated based upon this international theme of this guide. CEN/TC304 has a separate activity on European keyboard standardisation.

6.2 The processing function

The processing function provides for the manipulation of data according to the needs of an application.

Once input, the data is expressed in some internal computer system code. In addition, other information may be associated with each character such as colour, emphasis level and font. Such information is usually intended for some document processing function. Thus the system internal code structure may be quite complex. However, at its heart is the character code itself; document handling and processing is outside the scope of this guide.

Most commercially available computer systems do not use standardised character sets for internal representation of character data, but proprietary character sets or manufacturer specifications.

- The user needs to be able to have all input characters processed, while, again, the procurer needs to produce the appropriate procurement specification.

6.2.1 Ordering

A particularly common requirement on the processing function is that it be able to order character based data. The main ordering standard is currently under development as ISO/IEC 14651, *International string ordering – Method for comparing character strings and description of a default tailorable ordering*. ISO/IEC 9945-2 *POSIX Shell and utilities* also describes ordering. Standards for ordering, while related to character set standards, are not central to the theme of this guide. In CEN/TC 304, there is a separate activity on European standardisation of ordering.

6.2.2 character classification

Very important to character processing is the classification of characters as letters, digits, punctuation and control-characters etc. The main specification in ISO is ISO/IEC TR 14652, *Specification method for cultural conventions*, and ISO/IEC 9945-2 *POSIX Shell and Utilities* also has specifications on this. The cultural registry has data on character classification. Industry consortia like Unicode and X/Open have also specified data of character classification.

6.3 The interchange function

The interchange function provides for the interchange of data between computer systems. Since the character sets for processing are generally defined by manufacturer specifications, they are likely to be different in two different computer systems between which data are to be exchanged. Thus a character set for interchange is needed which will have to be different from one or both of the processing character sets. This is where character set standards become very relevant to reduce the number of interchange character sets – potentially one for every possible pair combination of different computer systems.

The main problem here is that there may not be a one-to-one correspondence between the characters in the character set for processing and those in the character set for interchange. In such cases less than trivial transformation functions are required at the interfaces between processing and interchange; see clause 10.2. Such transformations may incur a loss of

information.

- ▶ The user requirements will relate to the functionality of the interchange (e.g. what loss of information, if any, may be accepted). Such requirements may also include policy decisions of a more technical nature, e.g. as to what code structure to use. As before, the procurer's role is to transform the requirements into technical specifications.

A number of specifications exist for conversions between coded character sets. The main specifications are the charmap specifications contained in the ISO/IEC cultural register found at <http://www.dkuug.dk/cultreg/> - this is aligned with the Internet charset specifications. Also the ISO POSIX working group at <http://www.dkuug.dk/i18n/WG15-collection/charmaps> and the Unicode consortium at <http://www.unicode.org> have made data available for coded character set conversion.

6.4 The output function

The output function is the process of converting the internal coded representations of the characters to a visual representation on a display or hard copy device. The output character sets may be different depending of the output medium.

The handling of output to physical devices is usually an internal computer system function. Application programs, such as word processing packages, normally have the ability to control also the rendition of the output. This includes the use of fonts, both type and size, and also the use of various levels of emphasis and colour. In some cases, information which specifies particular values of these attributes is carried with the individual character codes right from the time of input. As already stated, these features are outside the scope of this guide.

The main problem is when the output character set is smaller than the character set for processing. The computer system software has to substitute one or more characters for those which cannot be represented by the output function (see clause 10.2). Again, such transformations may incur a loss of information.

▶ The user requirements will relate to the functionality of the output (e.g. what loss of information, if any, may be accepted). As before, the procurer's role is to transform the requirements into technical specifications.

6.5 Cultural issues

Each country or region in Europe (and elsewhere) has cultural conventions which affect the manner in which character sets are used by application programs. Such conventions include, but are not restricted to, ordering (already mentioned), numeric formatting, monetary formatting, date and time conventions, affirmative and negative answers, the use of special characters, personal name rules and classification of characters. The ISO/IEC cultural register is described and governed by ISO/IEC 15897 and the cultural registry may be found at <http://www.dkuug.dk/cultreg/>.

Cultural conventions primarily affect the processing function but may also impact the other functions. They are related to the use of character set standards but are not central to the theme of this guide. In CEN/TC304 there is a separate activity on cultural conventions which is developing and maintaining a registry of such conventions.

7 Official standards, manufacturer standards, and related standards

There are four main categories of character set standards:

- *Official standards for the computer and data communications industry*, promulgated by international standards organisations such as ISO/IEC and CEN, but also by national standards organisations. They are primarily intended for input (mainly for the specification of character repertoires) and for the interchange function in general applications.

- *Official standards for the telecommunications industry*, promulgated by ITU-T (formerly CCITT; the standards are called Recommendations) and ETSI, mainly intended for the interchange function in telematics applications.

- *Manufacturer standards for the computer and data communications industry*, developed by industry groupings such as the UNICODE consortium and individual companies such as IBM and Microsoft. They are mainly used for the input and processing functions but also for the interchange function in internal (proprietary) networks.
- *Related standards*, which involve character sets such as keyboard standards, ordering standards and transformation standards.

As already explained, this guide concentrates on the first of these categories (but does not cover national standards). In the remainder of this section further information is given on the other categories.

7.1 Telecommunication standards

Character set standards promulgated by ITU-T and by ETSI support applications primarily defined and promoted by the telecommunications companies. Thus for Teletex, Recommendation T.61 was developed. It was subsequently replaced by T.51, which is also used as the basis for character set support in other telematic applications such as the X.400 Message Handling Service and the X.500 Directory Service. There are some links, however, with ISO/IEC standards. T.51 is equivalent to ISO/IEC 6937 (see section 8); and there is also T.50, equivalent to ISO/IEC 646 (see section 8). This alignment is due to collaboration between the two standards organisations. Nevertheless, ISO/IEC 6937 is used primarily for telematic applications and services.

ETSI has developed character set standards in support of radio paging (ERMES), GSM and RDS, but these standards are not aligned at all with ISO/IEC standards.

7.2 Manufacturer standards

Manufacturer standards can be grouped into IBM EBCDIC variants and those in support of personal computers such as the PC and the Macintosh. The EBCDIC code pages have their own defined structure together with an invariant set of characters and code positions. This is similar in concept to the parts of ISO/IEC 8859 (see section 8). Indeed some EBCDIC code pages have the same repertoire as parts of ISO/IEC 8859, which makes it straightforward to perform mappings between the two environments. Other EBCDIC code pages are designed to support specific countries or regions. More information on this may be found in the IBM document "Character Data Representation Architecture" (see References).

Both IBM and Microsoft have designed PC code pages, and each page is aimed at supporting specific countries or regions. PC code pages do not in general comply with the 8-bit code structure specified in ISO/IEC 2022 (see section 8), and mappings between the two environments may result in the loss of information.

IBM EBCDIC and PC code pages are described in Appendix I of "IBM National Language Support Reference Manual Volume 2" (see References).

The document "Comparisons of Standardised Character Sets for Europe" provides mappings of the differences between various official and manufacturer standards (see References).

7.3 Related Standards

In addition to standards on the definition of character repertoires and coding, there are standards relating to the handling of character sets and other character set issues.

Keyboard standards are concerned with the input to computer systems of character information. They cover such topics as keyboard layouts. There is one international standard plus a range of national standards, most of which are based on the international standard.

Ordering standards are concerned with sorting

character strings or sets of strings into some order. There is an international standard, and work is under way in Europe on the ordering of European repertoires. A related activity concerns the matching of character strings, for use in, for instance, search engines.

Transformation standards are concerned with rules for mapping the characters from one repertoire onto another, and from one code system onto another. They are sometimes needed when character information from one computer system needs to be transferred into a different computer system due to a mismatch of functionality. An example is the mapping of an official standardised character set onto EBCDIC, so that character based information may be processed in an IBM machine. There are a number of different types of transformation needed for various scenarios. See also clause 10.2.

8 International character sets

This section describes internationally standardised character sets (i.e. both repertoires and coding) used primarily for input (mainly for the specification of repertoires) and interchange. For processing, many IT systems use proprietary manufacturer standards.

Because of the need to combine different standards and also in order to make the standardisation itself more consistent and effective, a common platform standardising the principles for code structure, code extension, implementation and registration has been established. The standards which define that platform may be called framework standards.

The international character set standards in this area may be classified in the following way:

- 7- and 8-bit framework standards
- 7- and 8-bit character set standards
- The universal character set (UCS) standard (multiple octet)
- The control function standard
- Character description standards

8.1 Framework standards for 7- and 8-bit environments

See also Figure 3 below.

- ISO/IEC 2022:1994, *Information technology – Character code structure and extension techniques*, is the basic framework standard. It provides the framework for the definition of all the 7- and 8-bit coded character sets and their use. In particular, it defines a complex code extension facility which allows considerably more characters to be used in a specific environment than would otherwise be the case with a single 7- or 8-bit code space. It also partitions the 7- and 8-bit code spaces between the representation of graphic characters and control characters. See also Annex A.

- ISO/IEC 4873:1991, *Information technology – ISO 8-bit code for information interchange – Structure and rules for implementation*, is a refinement of ISO/IEC 2022 which concentrates specifically on the implementation and use of the 8-bit coding. It lays the formal basis for the specification of two 8-bit character set standards – the various parts of ISO/IEC 8859 and ISO/IEC 10367 (see section 9).

- ISO/IEC ISP 12070:1996, *Information technology – International Standardised Profiles FCSnnn – Character set 8-bit code structure based on ISO/IEC 2022 – Part 1: FCS111 – 2022 Option 1*, is a profile of both ISO/IEC 2022:1994 and ISO/IEC 4873:1991. It goes into considerable detail concerning the use of character sets within Open System Interconnection (OSI) protocols and, in particular, concentrates on conformance requirements (in the interests of interoperability). It is particularly the latter which makes this standard relevant in this context, whereas its OSI aspects are less interesting nowadays.

Note – A *profile* is a selection of options from one or more standards, constituting a more narrow form of specification than the base standard(s) it refers to.

- ISO 2375:1985, *Data processing – Procedure for the registration of escape sequences*, governs the registration of coded

character sets. In the ISO/IEC 2022:1994 framework, the extension mechanisms use escape sequences to control the use of more than one registered coded character set. Such an escape sequence consists of the Escape control character followed by a variable number of octets. In this way, the sending implementation in an interchange can inform the receiving implementation about the elements of the code structure that will be used in the interchange.

A registration authority allocates values for the escape sequences to be used with the registered coded character set. The registration authority is the Information Processing Society of Japan/Information Technology Standards Commission of Japan (IPSI/ITSCJ), and the register may be accessed on-line at <http://www.itscj.ipsj.or.jp/ISO-IR/>. All standardised coded character sets are registered, but not all registered coded character sets are standardised. There are currently over 200 registrations. A coded character set may be identified in text, such as a procurement specification, by the sequence ISO-IR nnn, where nnn is the registration number.

<https://standards.iteh.ai/catalog/standards/sist/1aacdd30-5270-4299-ab08-f53cf3a09820/sist-tp-cr-13928-2003>

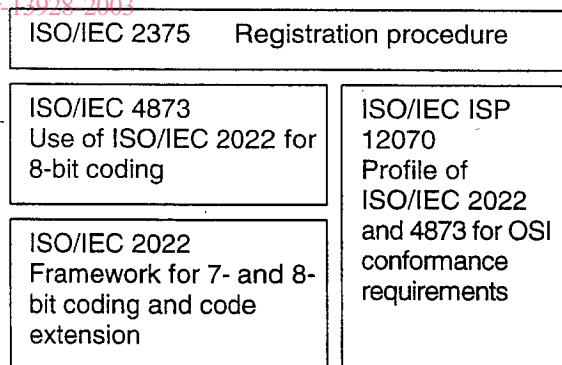


Figure 3 – Overview of framework standards.

► Most users and procurers need not concern themselves overmuch with ISO/IEC 2022:1994 or ISO/IEC 4873:1991, since their provisions are usually called up by the referring standards. For this reason, they need not be referred to directly in procurement specifications. However, the sophisticated user with a complex requirement, who may be mixing and matching the use of registered character sets