



NSAI
Standards

Standard Recommendation
S.R. CR 13928:2000

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

S.R. CR 13928:2000

Incorporating amendments/corrigenda issued since publication:

The National Standards Authority of Ireland (NSAI) produces the following categories of formal documents:

I.S. xxx: Irish Standard – national specification based on the consensus of an expert panel and subject to public consultation.

S.R. xxx: Standard Recommendation - recommendation based on the consensus of an expert panel and subject to public consultation.

SWiFT xxx: A rapidly developed recommendatory document based on the consensus of the participants of an NSAI workshop.

<i>This document replaces:</i> CR 13928:2000	<i>This document is based on:</i> CR 13928:2000	<i>Published:</i> 24 May, 2000
This document was published under the authority of the NSAI and comes into effect on: 14 May, 2011		ICS number: 35.040
NSAI 1 Swift Square, Northwood, Santry Dublin 9	T +353 1 807 3800 F +353 1 807 3838 E standards@nsai.ie W NSAI.ie	Sales: T +353 1 857 6730 F +353 1 857 6729 W standards.ie
Údarás um Chaighdeáin Náisiúnta na hÉireann		

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.



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.

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

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

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.

This is a free preview. Purchase the entire publication at the link below:

[Product Page](#)

-
- Looking for additional Standards? Visit Intertek Inform Infostore
 - Learn about LexConnect, All Jurisdictions, Standards referenced in Australian legislation
-