2000-10-31

SC22/WG20 N860

Draft CWA/ESR:2000

Cover page to be supplied.

Table of Contents

DRAFT CWA/ESR:2000 1

TABLE OF CONTENTS 2

FOREWORD 3

INTRODUCTION 4

1 SCOPE 5

2 REFERENCES 6

3 DEFINITIONS AND ABBREVIATIONS 6

4 GENERAL 7

5 ELEMENTS FOR THE CHECKLIST 8

5.1 Sub-areas 8

5.2 Characters 8

5.3 Use of special characters 10

5.4 Numbers, monetary amounts, letter written figures 11

5.5 Date and time 12

5.6 Telephone numbers and addresses, bank account numbers and personal identification 13

5.7 Units of measures 14

5.8 Mathematical symbols 14

5.9 Icons and symbols, meaning of colours 15

5.10 Man-machine interface and Culture related political and legal requirements 15

ANNEX A (NORMATIVE) 16

2000-10-31

FOREWORD

The production of this document which describes European culturally specific requirements on information and communications technologies was agreed by the CEN/ISSS Workshop European Culturally Specific ICT Requirements (WS-ESR) in the Workshop's Kick-Off meeting on 1998-11-23.

The document has been developed through the collaboration of a number of contributing partners in WS-ESR. WS-ESR representation gathers a wide mix of interests, coming from academia, public administrations, IT-suppliers, and other interested experts. The present CWA (CEN Workshop Agreement) has received the support of representatives of each of these sectors. A list of experts who have supported the document's contents may be obtained from the CEN/ISSS Secretariat. The final review/endorsement round of this CWA was started on 2000-02-17 and was successfully closed on 2000-mm-dd. The final text of this CWA was submitted to CEN for publication on 2000-mm-dd.

The CEN Workshop Agreement has only been made in English.

INTRODUCTION

Information and communications technologies (ICT) have been and continue to be undergoing breakthrough developments. Computers started as being helpful tools for highly repetitive or number crunching applications. For the sake of minimising the then exorbitant cost of processing cycles, computer memory (where each bit was still in the 70's physically represented by a magnetic ring with three wires running through it) and direct access storage, there was initially precious little room for variance and/or fine details at the system level in the coding schemes for entering and storing the data and at the application level for presenting the data. Although the technology base for the restrictions has long since gone, the user community at large is still in its infancy in the appreciation of what flexibility for cultural adaptability can be requested and expected from modern ICT solutions.

The system vendor community is busily building the infrastructure to meet the known and anticipated user requirements for this adaptability. While doing this, they face a number of challenging questions:

1. For which aspects/elements in the system is there a genuine justification to require cultural adaptability (and what are the actual conventions to be used in each case)?
2. How to provide this adaptability for use in application software (and how to make the application software houses utilise it)?
3. How to bring this adaptability to the end user in a transparent and seamless manner?

In line with the Protection of Cultural Diversity theme driven by the Commission of the European Communities (CEC), the primary intent of this CEN Workshop Agreement is to address the first part of question 1 above for Europe. In order to do so, the requirements of European nations and cultural groups need be covered, as well as those stemming from the natural interactions between these groups, in addition to those imposed by formally pan-European application environments.

The evident answer to address both parts of the second question is to use standardised methods, be they agreed upon within the formal International Organization for Standardization (ISO/IEC JTC 1, "Joint Technical Committee 1 for Information Technology" and particularly its SC 22/WG 20, "Programming Languages / Internationalization") or European Committee for Standardization (CEN ISSS, "Information Society Standardization System" and particularly its TC 304, "ICT - European Localization Requirements"), the World Wide Web Consortium (W3C, particularly its "User Interface: Internationalization"), the Internet Engineering Task Force (IETF), or whatever, as long as the forum is widely accepted. For any general solution, each IT system platform and each application must be "internationalised" in order for them to be able to be "localised". Unless the various platforms provide similar and ideally identical mechanisms (including the APIs, Application Programming Interfaces) for a broad enough set, the cost and other benefits of portability for cultural adaptability will not materialise. The answer to the second part also implies that the answer to the second part of question 1 is available.

The third question is the ultimate dilemma for the vendor community: A set of default values for each group of users must be easy to select (which is fully under the control of the vendors) and support their cultural conventions without modifications (which cannot be controlled by the vendors). These values themselves are not addressed by this CWA since, in order for them to be reliable, they must be provided at source, i.e. by the proper representatives of each nation and cultural group. Mechanisms are being developed for the registration of these values (e.g. the original European pre-standard ENV 12005 and its fast-tracked international version, ISO/IEC 15897:1999, which is currently under major revision) in an authoritative (yet hopefully neutral), reliable manner. In addition, for Europe, CEN TC 304 has set up a project team for the population of the cultural registry (as per ENV 12005) and another CEN/ISSS Workshop ("Eurolocale") is addressing the (ENV 12005 based) default values for use in pan-European applications. Unfortunately, the registries only cover a relatively small number of cultural elements and even as such are only sparsely populated. On the other hand, past, regional or proprietary efforts are known to contain a number of errors, and cannot thus be used without considerable caution.

When dealing with the default values for the cultural elements, it should be noted that once standardised methods are being used, the market relevance for any particular set of elements known to be correct does not require extensive proof. The addition of a table driven selection is a marginal investment (provided an existing user interface including

the keywords, error messages and help information may be utilised). In each and every instance, though, users must have the option to override any and all default values with their own personal, corporate or whatever preferences.

Culturally specific requirements are indeed culturally specific, i.e. the (established) cultural groups define them and can never be told what they are. Europe consists of several sovereign nations inside and outside the European Union. Even within the European Union, the strong subsidiarity principle guarantees that there can be hardly any pan-European conventions / cultural aspects that could be imposed on the national level, even for new elements such as the euro, the new currency of the European Monetary Union. As a consequence, no European standards (EN) can be expected in this area (since all CEN members are obliged to make each EN a national standard and withdraw any national standard that is in conflict with an EN). Truly pan-European applications, however, have a set of their own requirements to be met and European standards can thus be defined for them.

Note: *Thus, each nation and cultural group is strongly urged to:*
1. review that their inherent language and other cultural requirements are covered herein and
2. agree on one or more sets of the default values and submit them for registration.
This is an area, where one cannot depend on the others, except for the implementation once the homework has been done. Not doing this homework is tantamount to a public denial of the need for others to respect one's cultural background.

1 SCOPE

This CEN Workshop Agreement defines a check list of Culturally Specific ICT Requirements, such as character sets, internationalisation and user interfaces, in Europe (see Annex A for coverage) that products and services developed on the framework of the Global Information Infrastructure need to cover and support. Currently there is no single source for an integrated set of information regarding culturally specific ICT requirements in Europe. Such a checklist provided by the CWA will assist in this regard.

The CWA also discusses the rationale for the requirements that affect the localisation of ICT systems and services. In addition to the requirements in a national / cultural application environment, the CWA identifies areas where national requirements still need be addressed even in pan-European applications.

The potential users of this checklist are: 1) suppliers and implementers who wish to provide products and services applicable to the relevant market, and service providers, in particular those who wish to operate across national borders, and 2) users and purchasers who wish to ensure that products and services are applicable to their use.

The list, however, does not constitute any procurement guidance as such.

In areas, where reliable sources exist for the specifics of the requirements, the CWA will refer to these sources.

2 REFERENCES

EN 1923 European character repertoires and their coding -- 8-bit single byte coding
ENV 12005 IT -- Procedure for European Registration of Cultural Elements
ENV 13710:2000 IT -- European ordering rules - Ordering of characters from the Latin, Greek and Cyrillic scripts
CWA 13873:2000 IT -- Multilingual European Subsets in ISO/IEC 10646-1
CWA Eurolocale-1:2000 IT -- European generic locales -- Part 1: general specifications
CWA Eurolocale-2:2000 IT - European generic locales -- Part 2: narrative cultural specifications, POSIX locales and repertoiremap
ISO/IEC 646 IT -- ISO 7-bit coded character set for information interchange
ISO/IEC 2022 IT -- Character code structure and extension techniques
ISO/IEC 6937 IT -- Coded graphic character set for text communication -- Latin alphabet
ISO/IEC 8859-1 IT -- 8-bit single-byte coded graphic character sets – Part 1: Latin alphabet No. 1
ISO/IEC 8859-15 IT -- 8-bit single-byte coded graphic character sets -- Part 15: Latin alphabet No. 9
ISO/IEC 10646-1 IT -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane
ISO/IEC 14651:2000 IT -- International string ordering and comparison -- Method for comparing character strings and description of the common template tailorable ordering
ISO/IEC 15897 IT -- Procedures for registration of cultural elements
ISO 9 Information and documentation -- Transliteration of Cyrillic characters into Latin characters -- Slavic and non-Slavic languages
ISO 233 Documentation -- Transliteration of Arabic characters into Latin characters
ISO 233-2 Information and documentation -- Transliteration of Arabic characters into Latin characters -- Part 2: Arabic language -- Simplified transliteration
ISO 233-3 Information and documentation -- Transliteration of Arabic characters into Latin characters -- Part 3: Persian language -- Simplified transliteration
ISO 259 Documentation -- Transliteration of Hebrew characters into Latin characters
ISO 259-2 Information and documentation -- Transliteration of Hebrew characters into Latin characters -- Part 2: Simplified transliteration
ISO 843 Information and documentation – Conversion of Greek characters into Latin characters
ISO 4217 Codes for the representation of currencies and funds
ISO 8601 Data elements and interchange formats – Information interchange -- Representation of dates and times

3 DEFINITIONS AND ABBREVIATIONS

EDI Electronic Data Interchange
EU European Union
ICT Information and Communications Technologies
IT Information Technology
SWIFT Society for Worldwide Interbank Financial Telecommunications
UCS Universal Character Set
UN/EDIFACT United Nations Electronic Data Interchange For Administration, Commerce and Transport
UTF-8 UCS Transformation Format 8

4 GENERAL

Europe consists of many nations and the members of these nations speak different languages. Many nations are multilingual and virtually the same language is often also spoken by members of several nations. Furthermore, most European languages are related to some other European languages, although they form several groups. Although many "native Europeans" use languages such as Arabic, Gujarati, etc., to speak to members of their cultural communities, some of which have over 100 years of assimilation within Europe, these languages are not taken into particular consideration for the purposes of this CWA.

The sovereignty of the nations and the strong subsidiarity principle of the European Union have been discussed above. Furthermore, several European countries have already signed and ratified the Council of Europe treaty on regional and minority languages, which does imply certain responsibilities at the national level. More recently, the protection of cultural diversity has become an acknowledged goal for the Commission.

There is hardly any traditional cultural convention that is common to all of Europe. The early ICT implementations imposed a number of restrictions on these, and some of the less proprietary ones were even thought to have attained some level of permanence. Regarding the restrictions, this has now proved to be an illusion, of which an illustrative example is "Latin-1" (ISO 8859-1:1987), an 8-bit coded character set which supposedly met the requirements of also the French and Finnish languages (which was also at the time confirmed by the corresponding national standards bodies), whereas it is now understood that one needs to use the more recent "Latin-9" (ISO/IEC 8859-15:1999) in lieu of "Latin-1" to meet these requirements. Nevertheless, in spite of its hitherto repressive power, ICT has clearly also given rise to a tremendous number of very important, totally new cultural conventions of which many apply on a global basis.

To illustrate the problems, the languages of Europe have unique characteristics to such an extent that no single 8-bit coded character set can cover them all, not even those using the Latin script. Since the Universal Character Set (UCS - ISO/IEC 10646, also UNICODE), however, has yet to become the universally available base, e.g. common administrative procedures cannot be implemented in a straightforward manner honouring all cultural conventions even for the correct spelling of people's names. Since there are also other individual cultural conventions that cannot be followed in formal pan-European applications, these applications clearly need definitions of their own. Somewhat similar problems arise in less formal interactions already between multilingual individuals and they are heightened by both the implementation of the Inner Market and the advent of web-trade.

It should also be noted that even widely used symbolic representations of characters, words and other expressions are often highly culture and language dependent. For example, the Braille symbols, in both 6- and 8-dot versions, are language dependent and occasionally represent more than single characters. Thus, the encoding of the 8-dot Braille symbols in the UCS is only for the patterns and does not provide even a clue for their meaning. Another example of a similar cultural dependency is that of the sign languages for the deaf.

5 ELEMENTS FOR THE CHECKLIST

5.1 Sub-areas

In the following, each relevant sub-area begins with a table of entities that are known to vary within Europe, followed by a narrative text illustrating some of the variance.

5.2 Characters

For applications in a local cultural environment, normally based on the dominant language, i.e. not for pan-European applications.

Script	
Character properties	Some dependency on language.
Character repertoire	Dependent on (country and) language(s).
Transliteration rules	Dependent on (country and) languages.
Fallback schemes	Dependent on (country and) languages.
Transcription rules	Dependent on (country and) languages.
Ordering sequence	Dependent on (country and) dominant language.
Searching rules	Dependent on language.
Keyboard input	Dependent on country and language(s).
Optical character recognition	Dependent on language.

There are six modern scripts in use by native Europeans (Latin, Cyrillic, Greek, Armenian, Georgian, and Hebrew) and a number of other scripts are used by large immigration populations throughout Europe. The used script is often dependent on the language, although a given language may be written using more than one script. In addition, a number of both historic and artificial scripts have their own user communities. Graphic characters have a number of different, exclusive and non-exclusive properties (such as alphabetic, numeric or special; alphabetic characters may be of lower case or upper case or either one only; etc.) which may (in principle) vary from one language to another.

For practical purposes, the full character repertoire of a given language should be available for processing all at once (i.e. covered by a single coded character set), although it at times is overly difficult to agree on what constitutes the full repertoire, particularly for languages for which an established orthography has not been well documented. Even for widely well known languages, the question arises on

- which letters are essential to the language
- which letters are commonly used in writing the language, but not essential for the language
- which letters are used in names according to common practice, but not part of the language
- which letters are used seldom in writing the language
- which letters are “commonly“ used in writing foreign expressions

and thus where to draw the line for the full repertoire.

If the full required character repertoire (of one or more languages) is not available in a single coded character set, a number of schemes (e.g. the essentially exhaustive UCS short names, UTF-8 or ISO/IEC 2022, EN 1923 or ISO/IEC 6937 for text communication with a limited Latin alphabet, or some mnemonic scheme) may be used for the implementation.

Whereas the requirement above is to enable the use of the required letters, there is a parallel requirement to be able to limit the repertoire in a manner comprehensible to the intended audience. Thus e.g. Greek, Russian or Yiddish words or expressions (whether understandable or not) are likely to be readable by the majority of western Europeans only when written using the Latin script, and the same applies conversely to all scripts. Furthermore, even in the Latin script (or Cyrillic, particularly Slavic vs. non-Slavic), a number of characters with or even without diacritical marks (e.g. THORN < Þ,þ >) may be totally unfamiliar to a large number of native users of the script.

The limitation of the repertoire is usually done by transformation (and it may be enforced by not allowing other than certain codes in the system). Of the transformation methods, transliteration is a reversible process, which is often required for bibliography and other scholarly work. Transcription, on the other hand, is a process whereby the pronunciation of a given language is noted by the system of signs of a conversion language. Within the same script, irreversible, rather simple fallback schemes are often used (e.g. the removal of diacritical marks). The acceptability of such simple schemes strongly depends on the language.

Transliteration is a process which could conceivably be standardized between scripts, although it would not necessarily be very helpful to the layman. There are international standards for transliterating Greek and Cyrillic letters to the Latin script, but these standards are not widely known and deployed in everyday life. It is therefore often necessary to use transliteration based on national schemes at least as explanations to e.g. ISO 9, ISO 233 or ISO 259 conformant transliterations. Definite transliteration rules usually apply only from one specific language to another specific language and there are very few standards in this area. A case in point is e.g. how differently Russian names are properly written using the Latin script in e.g. English, French, German, Swedish or Finnish text (which is mostly due to the fact that in different languages several combinations of Latin characters are pronounced differently).

In most instances, the primary requirement is to be able to process together all the characters of a given language with the right character properties and the right ordering sequence (and searching rules) for the characters (and, eventually, combinations of characters). Normally the requirement is extended to being able to process together all the characters of all the languages that are commonly used in a given multilingual environment. With this extension, however, the processing rules must be common and follow those for the “primary“ language of the document and thus not those for the other languages. It should be noted that the same language in different countries may even have a different character repertoire and different processing rules. The processing rules, be they for their own character repertoire or fallback rules or whatever, cannot be imposed on the national level from the outside.

All characters of a given language should be enterable in a straightforward manner from a suitably adapted keyboard. This includes punctuation characters commonly used in the language according to its typographic practice, such as quotation marks and dashes. (A special case is Spain, where N WITH TILDE < Ñ,ñ >, INVERTED EXCLAMATION MARK < ¡ > and INVERTED QUESTION MARK < ¿ > must appear as such on the keyboard.) Furthermore, in a multilingual environment, one should be able to relatively easily enter single foreign characters or switch between different layouts for more extensive text. Any optical recognition of hand-written text must take into consideration the various ways of writing e.g. the numbers.

A special case is for the truly pan-European applications. As they are likely to be formally multilingual with no “primary“ language, their processing rules must be set specifically for them. For example, a list of names in a printed EU directory should always be in a predictable (i.e. the same) order and the same fallback schemes should always be followed. Thus, these processing rules (and the repertoire) could be standardised at the European level. In fact, several applicable repertoires have been defined by the MES Workshop in CWA 13873:2000. Furthermore, the pan-European ordering rules for the MES-2 repertoire have been defined in ENV 13710:2000, which serves also as the background specification for the ordering of whatever has not been specified locally, just as ISO/IEC 14651:2000 should be used to complement ENV 13710. For the purposes of this CWA, it is clear that the pan-European applications should have a set of their own values for the relevant cultural elements, some of which have been specified in CWA Eurolocale:2000. For viewing electronic documents, though, the application of language dependent or other personal preferences should be possible.

5.3 Use of special characters

The use of special characters for punctuation purposes varies considerably from one country to another. Differences are to be found particularly in the following areas:

Hyphenation	Dependent primarily on language.
Quotation marks	Dependent on country, language and usage: - left/right outer quotation marks, - left/right inner quotation marks, - left/right special purpose quotation marks.
Adjacent punctuation	Dependent on country (and language).
Ellipsis, dashes, spaces…	Dependent primarily on country.
Question marks, exclamation marks	Dependent on language.
Ditto marks	Dependent on country and sub-culture.
Check marks, other misc. symbols	Dependent on country and sub-culture.

The rules of hyphenation are totally language dependent and several schemes may apply even for the same language.

The use of the quotation marks is highly dependent on the country and the language. The quotation marks are included in the UCS, but whether one is in a given environment used as the left or the right quotation mark cannot be deduced from its thus misleading UCS name. Each proper pair of quotation marks (be they for outer or inner quotations) may be identical or they may be reversed or lowered or raised in whichever way, etc. Sometimes specific quotation marks may be used for legal purposes or to indicate an archaic expression. The order of adjacent punctuation is dependent on the language, which is highly evident with the quotation marks, where e.g. the final full stop is placed before the right quotation mark in English and after it in all Nordic languages (unless the quotation consists of the full sentence).

Similarly, one or more hyphens and dashes (including the minus sign) have some specific uses in some European countries. They may also be surrounded by some kind of spaces. The correct number of spaces (e.g. one or two after a FULL STOP) is also not fixed for all of Europe.

Question marks and exclamation marks usually follow the sentence, but in some languages the sentence will also be preceded by such an inverted mark.

Special symbols are often used to mark both correctness and an error (and a check mark in one culture can have the opposite meaning in another culture). Similarly, the ditto sign is often culture specific.

Various punctuation marks may or may not be used in a given cultural environment and some are used for a limited meaning only. Some examples of the more common ones with limited use include: In several environments, the NUMBER SIGN < # > is not used at all (but various other expressions, mostly some kind of abbreviations, are used instead). Similarly, the COMMERCIAL AT < @ > is in many environments replaced by < à > and as such only used for the addresses in the Internet and for other special usage, and the PILCROW < ¶ > is nowadays most commonly known only as an internal delimiter in text processing systems.

Thus, the proper use of all punctuation marks should be analysed for each culture (country/language, at the minimum those for which any automated corrective spelling and grammar checkers are made available) and appropriate means to better localise the IT systems accordingly at this level should be found.

5.4 Numbers, monetary amounts, letter written figures

For writing out the numbers, different schemes are used in the different countries of Europe.

Decimal separator	Dependent on country.
“Thousands“ separator	Dependent on country.
Monetary decimal separator	Dependent on country.
Monetary “thousands“ separator	Dependent on country.
Presentation of negative amounts	Dependent on country.
Presentation (+/-) with currency symbol	Dependent on country and currency.
Presentation (+/-) with currency code	Dependent on country.
Letter written figures	Dependent on country and language: - requirement, format, wording.
Ordinal numbers	Dependent on country and language.
Date	Dependent on country and language: - long form, - short form, - abbreviations.
Time	Dependent on country and usage.

In most countries, a COMMA < , > is used as the decimal separator, whereas in some countries (notably the United Kingdom and Ireland) a FULL STOP < . > is used for the same purpose.

For large numbers, readability is often enhanced by grouping the digits usually in groups of three. The “thousands“ separator is often some kind of a space (e.g. a NO-BREAK SPACE) or a FULL STOP < . >, but it may also be a COMMA < , > if the decimal separator is a FULL STOP < . >.

The monetary amounts have mostly similar decimal and thousands separators, although for certain currencies no decimals exist. The currency symbol may precede or follow the amount and it may or may not be separated from it by some kind of a space; there are also instances where the currency symbol is used as the decimal separator. A negative amount may be indicated by placing the minus sign before or after the amount, before or after the currency symbol. The presentation of amounts in euro may be different from those in the traditional national currency, particularly if the currency units are of different magnitude. The presentation of monetary amounts using the three-letter currency code (as per ISO 4217) may be similar to that with the currency symbol or different rules may apply; the numeric currency code is hardly ever used in text. For the purposes of this discussion, the currency symbol may be a sign (e.g. < £ > corresponding to the code ‘GBP’) or it may be a commonly used and often legal abbreviation of the name of the currency (e.g. < mk > corresponding to the code ‘FIM’).

There may be a (legal) requirement to spell out certain numbers also in letters, in which case specific rules may apply. It should be noted that written numbers may be different using the same language in different countries (e.g. French in France, Belgium, and Switzerland). Also, note that the word "billion" has become ambiguous in English, since its US meaning 'milliard' is in practice widespread in Great Britain, too.

Ordinal numbers may be identified by a period following the cardinal number, or the number may be followed by the ending (consisting of one or more letters) of the ordinal number, sometimes separated from the number by e.g. a COLON < : >. In certain instances, Roman numerals are used as ordinal numbers with or without a period, depending on the language.

5.5 Date and time

The official calendar is uniform in Europe as far as the days, months and years are concerned. The national holidays are different and several different church calendars exist, too. *– N.B.: Significant cultural minorities use other calendars and different year-end points, which may not have an one-to-one correspondence to European years. Nevertheless, dates from such calendars are admissible in evidence in European courts.*

The use of the international standard for the numeric date and time (ISO 8601) is in fact rather rare outside of specifically structured data for information interchange (as in EDI messages based on e.g. UN EDIFACT). Even in countries where this standard has been adopted, the use of other presentation forms is often even formally recommended (and its method of joining the time to the date is hardly ever seen). In the numeric form for the date, the sequence of fields may vary as well as their separator (often a FULL STOP <. > or a SLASH < / >) and leading zeroes may or may not be included.

This said, when a date is expressed using numerical designations only, the use of any notation that is not ISO 8601 conformant will very likely cause misunderstandings. The differences between European and US practices, as well as the differences in the placement of the year designation, make the use of two-digit years in any context highly undesirable; a notation like 01/02/03 has far too many possible interpretations, even if we assume that the year is not in the middle. Thus, depending on the context and audience, either ISO 8601 notation or a presentation containing the month in letters and the year in four-digit notation, or both in parallel, should be encouraged. Whether one of the alternatives is to be preferred depends on current national practice, but for pan-European purposes, the inclusion of the language-neutral ISO 8601 alternative is desirable whenever possible.

In the long form, the day may precede or follow the name of the month (which in most European languages does not start with a capital letter) and it may or may not be followed by a period. In several languages the more polite form requires the use of a particular expression, and in certain languages the name of the month must be inclined (and the inclination may be different depending on whether the month precedes or follows the day; e.g. in Finnish: ‘elokuu_n 12.’ or ’12. elokuu_ta’). The proper way of abbreviating the name of the month may have such restrictions that it becomes almost useless (e.g. in Finnish: ‘marrask.’ for ‘marraskuu’). In certain instances, e.g. the corresponding Roman numeral may be used in lieu of the abbreviated name of the month.

If the date includes the name of the weekday (which in most European languages does not start with a capital letter), this may have to be inclined (e.g. in Finnish: ‘torstai_na, 12. elokuu_ta’) or otherwise expressed in a specific way. The proper way of effectively abbreviating the names of the weekdays may encounter similar problems as found with the names of the months.

The separator for the fields of the time may vary (it usually is a COLON < : > or a FULL STOP <. >) and the leading zero for particularly the hours may or may not be omitted. Although the clock is a 24-hour clock for most official purposes, the 12-hour clock remains in wide use. The rules for spelling out the time are exceedingly variable.

5.6 Telephone numbers and addresses, bank account numbers and personal identification

Although there has been remarkable convergence in the presentation of telephone numbers and addresses, cultural differences still abound. The convergence, in fact, has mostly taken place in the markings required for international routing.

The common practice for expressing international telephone numbers is to precede the destination country code by a PLUS SIGN < + > to denote the source country dependent code for international dialling (of which there can nowadays be several in a number of European countries). The identification of the area codes and such and the grouping within the telephone number (and the separator used for that) is still highly country dependent (and not necessarily unified within a given country, either).

The rules for proper addressing, i.e. the use of titles and the way to write personal names (including the sequence), are highly dependent on the culture and on the nature of the correspondence. The legal implications are also to be considered (e.g. a letter with both the name of a person and the name of a company is in Finland in the extremes a personal letter if addressed to the person c/o the company and not at all personal if addressed to the company, to the attention of the person). In addition to stating the addressee for delivery, the formalities for addressing the recipient in the letter itself, i.e. the salutations, as well as the ending phrases are highly dependent on country, language and social custom.

The system of personal names is strongly culture-dependent. It generally contains a given name, but otherwise practices differ. In some cultures, surnames are used rarely if at all, whereas patronymics are in common use. In other cultures, patronymics are practically never used. The use of middle names, or middle initials, is partially culture dependent but within Europe it mostly depends on personal decisions. Both given names and surnames can be multi-part names, either as sequences of names, or names joined by hyphens, or names joined by other connectives such as the Spanish "y". Both the input and output of personal names as well as their presentation in files and data bases must be able to handle a large variety of alternatives, but in a specific cultural environment, the presentation often needs to be oriented towards a particular system, or other systems must be mapped to a particular system somehow. This is not only important for the practical identification of people; it is also a very sensitive area due to the personal nature and cultural significance of names.

The details of the delivery address also vary considerably. Thus, e.g. the house number, if any, may precede or follow the street name, and the identification of the staircase and the apartment, if any, may be combined with the house number or stated on a separate line.

The use of the postal code is converging in such a way that it will precede the name of the post office; the use of the country code preceding the postal code is less established, even for international mail.

The structure of the bank account numbers, including the delimiters and the check-sum algorithms used for the initial validation vary from country to country, although they are usually uniform within a country. For international money transfers, though, the SWIFT numbering scheme is used widely.

The use of the national personal identification numbers is often restricted to the extent that specific permissions must be obtained with proper justification to record them in any data base and to further display them in any document. The delimiters and check-sum algorithms vary from country to country even more than for the bank account numbers; the numbering scheme may also provide for the indication of the sex. The concept "personal identification number" is to be understood generically, since it need not be purely numerical. Similarly, algorithms and routines that read, check and process such identifications must not assume that only digits are acceptable. The repertoire of possible characters is to be specified for each country.

5.7 Units of measures

There is considerable movement in Europe towards the use of the metric system, but it is not yet uniformly adopted for general use by the population at large nor has it penetrated many of the specialty areas at all. One could argue that the differences in the units of measures do not necessarily affect the IT systems, but they do, particularly in e-commerce. In the ad hoc web trade, the customers may be totally unaware of even the possibility of the units of measures being different. For the more traditional units of measures (for e.g. the lengths, weights, etc.), the number of choices is rather limited, but for e.g. the shoe or dress sizes (with yet different sizes for men and women and boys and girls) or colour codes etc. the number of choices is overwhelming for a layman. Thus, some equivalency scheme must be provided, either an automatic one or, at the minimum, a lookup table (which is the solution currently implemented in postal order catalogues). Although one could argue that this is the responsibility of each seller, the uniform problem should have uniform solutions.

Units of measures, unless absolutely evident from the context, should be accompanied with information about the system of measures used. This implies that the format of either the data itself or the accompanying metadata must include provisions for including this information. This is particularly important in cultural contexts where Anglo-Saxon units might be used, especially since it needs to be made clear whether imperial or US units are used, and other unit qualifiers (e.g. for different types of ounces or barrels) need to be included.

Related to the problem with the units of measures, e-commerce also has a problem with the list of ingredients. If it is important for a customer to be aware of the possible presence of certain ingredients (for health, religious or whatever reason), he should be informed if not all the same ingredients that are mandatory to be listed in his own country have to be listed separately in a potential country of origin for the sale (provided that the customer has indicated the need to be informed of such differences in the legal requirements). The ingredients are often listed, due to space constraints and other practical considerations, so that the list in one language lists them according to the regulations of the country where that language is dominant; but especially from the viewpoint of language minorities, such an approach cannot be regarded as generally acceptable, at least not without explicit caveats. To complicate the issue, the naming conventions, including the E numbering scheme, are not necessarily uniform.

5.8 Mathematical symbols

Several mathematical notations are in use by the experts, depending more on their area of expertise than their country of origin, although the latter also has some impact on the choice.

For use in general text, the PLUS SIGN < + > and the MINUS SIGN < - > are in general use for addition and subtraction, although other signs may be used, particularly for subtraction (including the DIVISION SIGN < ÷ >). For multiplication, several signs may be used, e.g. MULTIPLICATION SIGN < × >, MIDDLE DOT < • >, or ASTERISK < * >. Similarly for division, several signs may be used, e.g. DIVISION SIGN < ÷ >, SLASH < / >, COLON < : > or the division line.

5.9 Icons and symbols, meaning of colours

Symbols and colours can have different meanings for a variety of reasons, although variation according to nationality or language is smaller than between cultural subgroups and individuals. As a result, even the established Red Cross/Crescent is currently in the process of defining a new, universally acceptable symbol.

A particular example of cultural dependency is the expanded use of North American traffic signs as general symbols. Such use is seen as offensive colonialism by some, and from the communicative viewpoint it unnecessarily presumes acquaintance with a specific system of signs.

The use of national flags to denote language, although convenient to implement, is another potential source of major irritation to the users, particularly for native speakers of the language outside the thus indicated country.

As a general rule, any use of icons and colours should be accompanied with a possibility of accessing textual explanations. This is important for accessibility reasons, including the needs of visually impaired people, but also because of the possibility of unclear or ambiguous meanings of icons and colours in different cultures.

5.10 Man-machine interface and Culture related political and legal requirements

The man-machine interface aspects, originally intended to be expressly covered by this CWA, are only alluded to in some of the previous sections. As regards the various modes of data capture, in particular, culture related specific requirements do exist but they would be presently rather difficult to formulate meaningfully for the emerging technologies.

As an example of culture related interface cases, the keyboard problems require consideration as separate from character availability issues. Even if a character needed for man-machine interaction can be expected to be widely available, it can be difficult to produce on different keyboards. Any character outside the invariant part of ISO 646 may cause problems for such reasons. Consequently, in the design of interfaces, it is desirable to have alternative notations, and country-specific studies in the properties of commonly used keyboards might be useful to aid the design of interfaces.

A variety of culture related political and legal requirements exist, although the details thereof are seen to be outside the scope of this CWA. Some of them (e.g. which languages must be supported for any given purpose) stem from the various national interpretations of the consequences of signing and ratifying the Council of Europe Charter for Regional or Minority Languages, or e.g. the Northern Ireland peace deal. Other requirements deal mostly with the application level and only very few with the basic hardware (except for safety) and software implementations. Some examples, though, are given in the previous sections. The importance and use of mandatory procurement conditions is being reduced.

An important source for requirements is the mandated standards with the “new approach“ directives, i.e. the Community laws, where the technical details of the regulations are not included in the directive but expected to be defined by the relevant standards making bodies. These are being widely introduced in the telecommunications area. New, emerging areas are health care and electronic signatures.

Annex A (Normative)

The list of countries covered by this CWA on European Culturally Specific ICT Requirements is based on them actually having given CEN some specific authority to act for them in the area of ICT standardisation. They are the Members of EU ('E'), EFTA ('F') and/or CEN ('C'), plus CEN Affiliates, i.e.:

Members: Austria (E,C), Belgium (E,C), Czech Republic (C), Denmark (E,C), Finland (E,C), France (E,C), Germany (E,C), Greece (E,C), Iceland (F,C), Ireland (E,C), Italy (E,C), Liechtenstein (F), Luxembourg (E,C), Netherlands (E,C), Norway (F,C), Portugal (E,C), Spain (E,C), Sweden (E,C), Switzerland (F,C), United Kingdom (E,C);
CEN Affiliates: Albania, Bulgaria, Croatia, Cyprus, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, Slovenia, Turkey.

The CWA may also allude to information on other countries.