ISO/IEC JTC 1
Information Technology

ISO/IEC JTC 1 N 5449

DATE: 1998-06-24

REPLACES

DOC TYPE:
National Body Contribution

TITLE:
Contribution from The Netherlands on the Functioning of ISO/IEC JTC
1/SC 2, Coded Character Sets

SOURCE:
National Body of The Netherlands

PROJECT:

STATUS:
The JTC 1 Chairman will discuss the issues raised in this paper with
the SC 2 Chairman and will advise JTC 1 of the results of that
discussion.

ACTION ID: ACT

DUE DATE:

DISTRIBUTION: P and L Members

MEDIUM:

DISKETTE NO.:

NO. OF PAGES: 10

Secretariat, ISO/IEC JTC 1, American National Standards Institute, 11
West 42nd Street, New York, NY 10036; Telephone: 1 212 642 4932;
Facsimile: 1 212 840 2298; Email: lrajchel@ansi.org

Title: CONTRIBUTION FROM THE NETHERLANDS ON THE FUNCTIONING OF
                ISO/IEC JTC1/SC2, Coded Character Sets

Source: NNI, The Netherlands

Status: To the attention of ISO/IEC JTC1

Date: 1998-05-25, VERSION 1.2

It is with considerable hesitation that we decided to take the step to
ask for your attention to an issue we think most important. The
Netherlands NB has become increasingly concerned about the way JTC1/SC2,
Coded Character Sets is functioning. In particular we want to highlight
three aspects of the first interest to us.

-- Stability of ISO standards is of primary importance to bodies, users
or industry, that have implemented these. The fundament of the prestige
of ISO is that the documents it provides will not be modified without
proper justification and only after careful consideration of ensuing
consequences. On this solid base industry produces implementations, and
users invest their money in procuring these. ISO (and IEC) TCs, SCs and
WGs have the duty to respect the confidence users have put in the
products claiming conformance as they are presented to them.

SC2 has not paid attention to serious objections to some proposals that,
if adopted, would endanger substantial investments. The extent of these
is illustrated in Attachment A. A particularly important case is
described in Attachment B.

-- Resources of NBs are quite limited and experts scarce. Thus JTC1
decided to stress market relevance with ISO standards to be developed,
in order to prevent issuing standards only relevant to small groups.

SC2 decided to disregard a request from our NB (SC2 N 2881, Attachment
C) to create rational limitations to additions of new characters or
scripts to SC2 standards. Its WG2 adopted instead a "roadmap" (SC2/WG2 N
1499), for ISO/IEC 10646, that even provides for inclusion of scripts
not yet decyphered (like that of the unique Phaistos disc) next to those
as esoteric as Etruscan, Sinaitic, Avestan, Old Persian cuneiform,
Buginese, Old Mormon and the scripts from the books of Tolkien.
Obviously, SC2/WG2 expects that NBs are prepared to send experts to
discuss these subjects a full week twice a year, an excessive claim to
NBs not having an interest in the scripts, but not wanting to be
confronted with unannounced proposals for undesired additions to their
national repertoire. These NBs require a standard like a safe house, not
like Ye Olde Curiosity Shoppe. After Amendment 7 to ISO/IEC 10646-1
industry expressed its concern about the growing numbers of these
(SC2/WG2 N 1388). But at present SC2/WG2 has arrived at PDAM 22.

-- Clarity of the decision process has always been an important goal of
ISO, in order to achieve broad support and consensus on precise wording.
Thus a system of steps and stages was meticulously maintained, resulting
in documents of which the detailed development could be well traced,
preventing any surreptious action. JTC1 has its Directives to rule the
process. Decisions may be taken by letter ballot or at meetings. But
proposals not yet documented may turn up at the latter, and may be
approved by delegates attending, without proper investigation of
consequences. We want to ask for attention to this problem, and want
guidance from JTC1 what to do if unwanted effects appear to have been
overlooked.

SC2 faced the problem several times. At the 1992 SC2/WG2 meeting the disposition of comments to DIS 10646-1 resulted in characters added, to which some NBs would have objected, had they known the attempt. Under heavy pressure delegates agreed to a coding scheme for Korean, which proved to be utterly impractical, and had to be replaced a few years later, after action from Korea, by a better scheme.

DIS 8859, parts 1-6,9,10, were out for letter ballot, ending 1997-07-06. SC2/WG3, meeting 1997-07-04,07, nevertheless discussed the contents of these parts and agreed to technical changes to the texts (SC2 N 2933, Resolution M12.03). At this moment the result of voting was not known, nor the NB comments. A resolution was adopted to apply a technical change to the 8 DISs. Preparation of the D of C and the final text was delegated to the Project Editors. After some time it was discovered that changes would make the 8 parts of 8859 inconsistent with ISO/IEC 4873, on which these are based. Also the layout of Table 2, the code table, in the 8 parts would be changed (but this decision is not found in any resolution). But ISO 2375 by which code tables are registered requires that these are presented in the style of ISO/IEC 646 or 4873. This was reported by our delegate (also project editor of 6 parts) to the next SC2 plenary meeting March 1998 in Seattle, USA. The fact could not be denied, but, instead of reconsidering the issue, it was decided to change ISO 2375 and to instruct the Registration Authority to modify their Practice document (SC2 N 3077, Resolution M08.06). Can a JTC1 SC instruct a JTC1 Registration Authority?

It should be realized that the technical changes in DIS 8859 only pertain to informative additions, but that the SC2 decision involves revision of all 200 registrations, and beyond that, of 646, 2022, 2375, 4873, 6937, 10367, and the other parts of 8859. And this is done without any cost/benefit analysis. No letter ballot was initiated to support this far-reaching action, and NBs not represented in Seattle may not even be aware of what is going to happen, and what it will cost them.

In the preceding we we have outlined our concern on the major points, leaving minor ones out (like the introduction of a new type of editor, Resolution M08.24).

As a conclusion we submit the following requests to JTC1.

1. We request JTC1 to check the present functioning of its SC2, in particular the disrespect shown to user needs, and to restore regular practice, not in the least by enforcing SC2 to keep to JTC1 Directives.

2. We request JTC1 to examine critically the program of work of SC2 and to remove everything serving only a small audience.

3. We request JTC1 to instruct SC2 to properly address the issue forwarded in the Position Statement of our NB (contained in Attachment B).

4. We request JTC1 to investigate the validity of SC2 decisions with respect to technical changes to the approved DISs 8859, parts 1-6,9,10, and to suspend publication of any part until the result is known.

Attachments
A National Activity Report to SC2 plenary, March 1998 (SC2 N 3046)
B Position statement of the Netherlands NB on the separation of
   characters
C Position statement of the Netherlands NB regarding further
   development in JTC1/SC2 (SC2 N 2881)

ATTACHMENT A

NATIONAL ACTIVITY REPORT FROM THE NETHERLANDS NATIONAL BODY
TO ISO/IEC JTC1/SC2 Plenary, 1998

The responsible committee for JTC1/SC2 matters in the Netherlands is
NNI 381 02.

Meetings are held now at irregular times only. Agreement on votes is
reached by e-mail.

INTERNATIONAL

On the international level the participation in TC 304 of CEN is being
restored to normal, but still restricted level, and contribution to
projects is under consideration. We hope that TC304 is returning now to
a realistic view of the world.

NATIONAL

A "POSIX LOCALE" for Dutch is under development, but halted due to
lack of time of the team members.

The revision of the national standards for handling personal data
(NEN 1888) and adresses (NEN 5825) is now under way. There are in
fact two projects.

1. Revision of:
   NEN 1888, General Personal Data
   NEN 5825, Addresses

   This is prepared by NNI NC 380 007
   Represented are:
   Ministry of the Interior
   Ministry of Justice
   Ministry of Finance (Taxation)
   Police Information Centre
   National Chipcard Platform
   Ass. of Neth. Communities
   Social Security
   Ediforum
   GBA (Communities Data Exchange)
   PTT Post Media service
   State Agency of Road Transport
   Health and Welfare Informatics
   Health and Welfare Insurance
   Housing Corporations
   Broadcasting Subventions
   Direct Marketing Ass.
   Some Service bureaus, among them:
   Human Inference
   Directview
   (some I left out for which I did not know the English equivalent)

A subgroup is working on character set matters.
This work covers:
-- selection of coded character set standards to be used,
-- conversion between Latin character repertoires,
-- transliteration of non-Latin characters to Latin.

2. Police Information Centre
   To select character sets for use with Police systems, and their
   mutual conversion.

Both projects work in close cooperation.

There are no decisions as yet, but the directions are already clear.
These may be summarized as stated in the following document:

                                                      VERSION 0.3
                                                       1998-02-17
                                                  J. W. van Wingen

Principal trends in character handling in systems of the Netherlands
Government or related institutions

1. It cannot be expected, nor required, from a civil servant to be able
to handle non-Latin scripts. Thus handling of these scripts will not be
included in the normative parts of the standards under development.

2. What matters primarily for data exchange and storing is the
repertoire of characters, not their coding.

3. In principle three repertoires will be specified for use:
        a small one,
        a middle one,
        a large one.
The selection will be directly related to the levels of Edifact (ISO
9735).

4. The small repertoire may be ASCII (ISO/IEC 646 IRV:1991) or Edifact
Level B (invariant subset of ISO/IEC 646:1991, ISO-IR 170).

The middle repertoire will be that of ISO/IEC 8859-1 or 8859-9. (Whether
there is, or will be, any actual use of it is a matter of doubt.)

The large repertoire is the GBA set, that is that of Teletex (T.61 or
T.51, subset of ISO/IEC 6937), without the IJ as a single character.
(This is part of the specifications for the GBA system which forms the
basis for all personal data communication and for authentication in the
Netherlands.) No additional characters will be permitted.

5. ISO/IEC 10646-1:1993 is considered not mature nor stable enough to be
recommended, apart from a carefully selected subset, which is identical
to the GBA set. It may offer coding for the large repertoire, but there
is no intention to change the teletex coding with the GBA system to full
two-octet coding.

6. For the coding of the GBA set three methods are considered suitable.
   a. that of ISO/IEC 6937
   b. that of UCS-2
   c. that of UTF-8
The choice will depend on the application (database, network).

From the trends it should be clear to SC2 that there is no interest at

all for promoting larger subsets than the GBA set for use with any body related to the Government. Even the interest for transliteration rules from other than Latin scripts is very restricted. The Committee is only permitted to discuss that if time allows.

It should be realized that not keeping to the principle of Latin script could even involve the risk of conflicts with the Unions. Thus there is no demand, and thus no market, for any large scale implementation of other scripts in the Netherlands, apart from a very limited production of Arabic texts for Moroccans living here.

It is under consideration to make the GBA set a EN from CEN. This may take some time, because the specification of the GBA set (taken as a repertoire or as a subset of 10646-1) is under review at present. It has been discovered that there are subtle differences between ISO/IEC 6937, ITU T.51 and the former ITU T.61 (Teletex). These pertain only to special characters, not letters. For letters the GBA set has the same repertoire as ISO/IEC 6937, and that is fixed.

Because this set of letters (subset of the GBA set) is also in use with ITU, and with X500, it will become in the future a very important thing for the whole of Europe, in particular because it has been stable since 1983 at least. This may disappoint some people, but the cost of change will be very high, and as has been decided, will have to be borne by the proposers.

GENERAL

The Netherlands National Body has noted with concern the way NL proposals were treated at the Crete meetings. In particular, the manner in which our statements were received where we exposed the risks to the stability of our administrative systems and even to national security, if certain additions to our character repertoire would have to be adopted, can only be called frivolous.

SC2 should realize that changes like these do not have technical aspects only, but have severe political implications. Should SC2 continue its course, then it should not be surprised when it would become involved in a diplomatic conflict between nations.

ATTACHMENT B

POSITION STATEMENT OF THE NETHERLANDS NATIONAL BODY
ON THE SEPARATION OF CHARACTERS

J. W. van Wingen   1998-03-10   version 1.2

Dear Colleagues

At the SC2/WG2 meeting, July 1997, in Crete the position was taken that:

LATIN SMALL CHARACTER S WITH CEDILLA
LATIN SMALL CHARACTER T WITH CEDILLA
would be different from:
LATIN SMALL CHARACTER S WITH COMMA BELOW
LATIN SMALL CHARACTER T WITH COMMA BELOW

The Netherlands delegate did not get the chance to submit a paper
explaining why we were so much against. Thus our arguments were not
sufficiently listened to.

Since that meeting we have investigated what the consequences would be
for our systems and those elsewhere. And these are so serious that the
Netherlands is requesting SC2 to withdraw any resolution that supports
the separation of the characters in question. Furthermore, it was
discovered that facts presented about these were not based on reality.

1. Romanian
We studied many papers in Romanian, in particular with the valuable
help of the library of the Institute for Eastern European Law of Leiden
University, which has more than 300 books in Romanian. It became clear
that many documents use cedillas in some font, and comma below in some
other, often on the same page (most recent of 1996), without making any
difference in meaning.

2. Turkish
A first look into a Turkish newspaper (bought last Monday) showed many
letters S with a comma below, obviously meant as a cedilla, and real
cedillas elsewhere on the front page, dependent on the font used. This
means that also in Turkish no difference in meaning exists, and thus
that there is no reason to assign a separate code for each. Still worse,
suppose that this page has to be coded, which code should be chosen for
a letter? This situation would create much confusion in Turkey.

The conclusion is that it not possible to distinguish in a mixed text
Romanian / Turkish what is what. Not even the use of a magnifying glass
would be of any help.

3. Exchange of personal data
In the free traffic of persons over the world it is important that they
can be identified with their correctly spelled name. If a Romanian wants
a permit to stay in our country, his name has to be entered into our
personal registration system (GBA). This is based on Teletex, subset of
ISO/IEC 6937, or ITU T.51. Because GBA has been established by law, its
conventions are hard to change, not to speak of implementations already
working. But not only money is involved, should an extra letter be
added. The new letter should be easily distinguishable from the others.
This is not the case with cedilla vs. comma below. Thus the Romanian
will be told that his name will be entered with a cedilla, and that he
has to sign that he agrees to that. If he does not, he will not get a

permit, and will not be allowed to stay.

Suppose that he does not accept this decision, and goes to complain to the National Ombudsman, or worse, to the European Court of Justice. Then it will be important in the case that follows that an ISO standard supports making difference between cedilla and comma below. Should the Kingdom of the Netherlands for that reason loose its case, the damage will be enormous, and may lead to destabilisation of our administrative system.

More is affected. The Police also uses GBA for registering criminals. Introducing different spellings could cause more confusion that we are prepared to accept. We would rightly claim that National Security is at stake.

Under these circumstances it will not surprise anyone that we are determined to oppose the separation of these characters with all means available. Should SC2 not come to its senses, we'll appeal to JTC1, and higher if needed. I appeal to you, dear colleagues, to reconsider the matter, in order that we arrive at decisions that are in the public interest.

ATTACHMENT C

POSITION OF THE NETHERLANDS NATIONAL BODY (NNI)
REGARDING FURTHER DEVELOPMENT IN JTC1/SC2

Now that JTC1 Reengineering is on its way, it is a good moment to
consider what should be the policy of the Netherlands regarding further
development of standards by JTC1/SC2. The primary points of concern are
future parts of ISO/IEC 8859 and new amendments to ISO/IEC 10646-1, but
some general aspects should be considered first.

General considerations

First of all, the stated principles of JTC1 should be taken very
seriously. Market-relevance should guide selection of projects. This
does not mean that academic preferences should be ignored, only that
standards institutes, depending on industry contributions, cannot be
expected to subsidize academic research. If Learned Societies want to
raise their agreed conventions to the status of an International
Standard, they should take the way of a Fast Track procedure, after
having done the development themselves.

The criteria of JTC1 are rather vague, but may be made more precise
for SC2. The actual expected use of a standard for coding a script
should be made clear on base of verifiable figures of published books.
and periodicals and the use in schools and courses.

The requirement of 5 participating NBs in a new project may be too
strict for SC2. It may mean in practice that single-nation scripts
cannot be included in an ISO standard. To acquire 4 other NBs may
involve dealings like "if you participate in my project on a script you
do not understand, I'll do in yours which I do not understand either".
Always saying YES as a result implies in fact loosening the criteria for
participation, because an unqualified YES does not raise questions.

On the other hand, no proposal for coding a script shall be approved
without a formal declaration on government level, or from a learned
society, that the proposal suits their needs.

+++
Future parts of ISO/IEC 8859.

From several sides concerns were expressed about so many proposed new
parts. We should not disregard these.

Some say that all coded data should transmigrate to data coded with UCS.
They ignore the market. As long as any accented letter will be called a
"foreign character" in the US, and 7-bit mailing systems are permitted
to exist, use of ASCII will continue. Thus the future will be that of
coexistence (and I hope a peaceful one) between single-byte and
multiple-byte coding.

To provide single-octet coding for characters not yet covered by
existing parts of 8859, two ways are open, that of Registration of a 94
or 96 character code table, suitable for a Right half (G1) according to
ISO/IEC 4873, and that of a New Part of 8859.

Registration (ISO 2375) was meant to identify a code table for
referencing in telecommunication and applications of ISO/IEC 2022. It
provides a Number (ISO-IR xxx) and a Final Byte. Any Sponsoring Body can

apply for registration of a code table. SC2 may comment on the application, but cannot change or reject it, unless it does not conform to the rules of ISO 2375. Thus further extension of the number of registered code tables cannot be stopped by any means.

If a code table is considered of more than restricted importance, and conformance to a standard is required for legal contracts, a New Part of ISO/IEC 8859 may be proposed. Then SC2 can accept the proposal or not, and it can change the code table. To complete the standard, a Final Byte must be known, which is assigned at Registration. Thus, registration is a de facto requisite for having a new part off 8859 published.

The practice with newly developed parts of 8859 is that some appear to be used little or not at all. So we have Part 4, which should be superseded by Part 10, which in its turn was rejected by the Baltic Countries who proposed Part 13. We cannot continue this way. We in SC2 should be sure that a proposed code table is already in actual use, and responds to practical demands. For Part 13 we got assurance from Latvia.

At present there is ISO-IR 182 for Welsh, with a code table in which some special characters are replaced by letters. There is a new part of 8859 proposed for Celtic, including Welsh and replacing more special characters with letters (for Irish). How do we know what the Welsh prefer: Irish letters (which they do not use), or more special characters? (That preference for less letters was one of the reasons why the Baltics rejected Part 10.) At present we are not convinced that the part for Celtic addresses a sufficient market.

This brings us to a:

DRAFT RESOLUTION TO SC2: Policy regarding new Parts of ISO/IEC 8859

1. A new code table meant for future standardization as a new part of
    ISO/IEC 8859 shall be forwarded first for registration according to
    ISO 2375 to the Registration Authority.

2. If after some period, not shorter then two years, it can be
    convincingly demonstrated to SC2, based on reports from NBs and
    industry, that the code table is in actual use to a sufficient
    extent, then SC2 forwards a NP for a new part of ISO/IEC 8859 to
    JTC1.

To our opinion the code tables for Latin/Thai, Latin/Devanagari (derived from ISCII) and Latin Alphabet no. 7 (Baltic Rim) satisfy the actual use requirement just now. We are not yet convinced that Latin Alphabet no. 8 (Celtic) does.

We request a decision from SC2 on this policy by resolution on its next plenary meeting.

+++
Further extension of ISO/IEC 10646.

Before any further additions to ISO/IEC 10646 are being made, we require that a priority policy be established on which scripts are candidates for later inclusion in the BMP. We want to have a decision now, before any irreversible decision is taken, now that only 10336 positions are left. Accepting new scripts on a "first in" base may cause exclusion from the BMP of some important scripts for which at present no proposal for coding has been submitted.

We consider the present classification (A,B,C, etc.) unsatisfactory. In particular, category A, "contemporary", contains very disparate

elements. We request that it be split into:

AA scripts for official languages of a state or important region, used
for publication of:
-- all documents of that state or region,
-- at least one daily newspaper,
-- numerous books.

AB scripts for regional languages as currently written, used
for publication of:
-- some official documents,
-- at least one weekly paper,
-- new books at regular times.

AC additions of a few letters to better suit the usage, at most 10.

When assigning characters to positions, category AA should always have
top priority. Even if no proposal exists, room should be kept free.
Scripts in this category not yet included in the BMP are:
Mongolian, Singhalese, Burmese and Khmer, possibly also Maldivian. We
are not aware of others. An estimate should be made of the number of
positions needed for those AA languages. Should this not exceed 10336 –
6656 = 3680, then we will not object any longer to the IRG
recommendation.

We are, in principle, not in favour of including in the BMP of new
scripts of regional or historical nature. These may be coded in other
planes of 10646.

This brings us to a:

DRAFT RESOLUTION TO SC2: Policy regarding new Amendments to ISO/IEC 10646-1.

1. Any script not yet coded in ISO/IEC 10646-1 shall only be included
    for coding (that is in the BMP) if it is convincingly demonstrated
    to JTC1/SC2 that it is in use as an official script for one or more
    languages of a certain state (member of the United Nations).

2. Before voting on a PDAM, a declaration issued by the Government
    of that state or by an Institution of sufficient authority dealing
    with the script, that the proposal suits their needs, shall be
    delivered to SC2 secretariat, which will circulate it to SC2
    membership.

3. Any extension of an existing script to be included for coding, shall
    be subjected to investigations by SC2 whether these characters are
    in use frequently enough to be of market relevance.