

ISO
 INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
 ORGANISATION INTERNATIONALE DE NORMALISATION
 МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ

ISO/IEC JTC1/SC2/WG2
 Multi-Octet Coded Character Set

ISO/IEC JTC1/SC2/WG2 N **1813**

Date: 1997-06-09

This is an unofficial HTML version of a document submitted to WG2.

Title: Proposal to add 10 Cyrillic Sámi characters to ISO/IEC 10646

Source: Trond Trosterud, Barentssekretariat (NO)

Status: NTS, Norwegian Member Body Contribution

Action: For consideration by JTC1/SC2/WG2

This document contains the proposal summary (ISO/IEC JTC1/SC2/WG2 form N1352) and a full proposal for the encoding of 10 Cyrillic characters in ISO/IEC 10646.

A. Administrative

1. Title	10 Cyrillic characters for Kildin Sámi
2. Requester's name	<u>Trond Trosterud</u>
3. Requester type	Member body contribution
4. Submission date	1997-06-09
5. Requester's reference	http://www.indigo.ie/egt/standards/se/kild.html , ISO-IR 200
6a. Completion	This is a complete proposal.
6b. More information to be provided?	No

B. Technical -- General

1a. New script? Name?	No
1b. Addition of characters to existing block? Name?	Yes, Cyrillic
2. Number of characters	10
3. Proposed category	Category A
4. Proposed level of implementation and rationale	Level 1; see Appendix A
5a. Character names included in proposal?	Yes
5b. Character names in accordance with guidelines?	Yes
5c. Character shapes reviewable?	Yes (see Appendix A)
6a. Who will provide computerized font?	Michael Everson, Everson Gunn Teoranta
6b. Font currently available?	Michael Everson, Everson Gunn Teoranta
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes (see Appendix A)
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes (see Appendix B)
8. Does the proposal address other aspects of character data processing?	No

C. Technical -- Justification

1. Has this proposal been submitted before? Explain	No
2. Contact with the user community?	Yes, with Saamskij sektor, Akademia NAUK, Murmansk
3. Information on the user community?	Limited (see Appendix A)
4a. The context of use for the proposed characters?	Common
4b. Reference	Appendix A
5a. Proposed characters in current use?	Yes
5b. Where?	In the Kola peninsula.
6a. Characters should be encoded entirely in BMP?	Yes
6b. Rationale	All Cyrillic characters should be in the BMP
7. Should characters be kept in a continuous range?	No, but they should be kept together with the other Cyrillic characters.
8a. Can the characters be considered a presentation form of an existing character or character sequence?	Yes, for 2 of the 10 characters
8b. Where?	
8c. Reference	See Appendix A
9a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?	No
9b. Where?	
9c. Reference	
10a. Combining characters or use of composite sequences	No

included?	
10b. List of composite sequences and their corresponding glyph images provided?	No
11. Characters with any special properties such as control function, etc. included?	No

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC 2/WG 2 document numbers:	
2. Status (list of meeting number and corresponding action or disposition)	
3. Additional contact to user communities, liaison organizations etc.	
4. Assigned category and assigned priority/time frame	
Other Comments	

E. Proposal

Inclusion of 10 character positions for Kildin Sámi in ISO/IEC 10646

Trond Trosterud, Committee for Character Set Technology, Norsk Teknologistandardisering

Historical background for the Kildin Sámi script

The development of literary Kildin Sámi follows the path of the other non-Slavic languages quite closely. It was written for the first time in the second half of the last century, in form of religious texts based on the Cyrillic alphabet. In the late twenties and early thirties the Institute of the Northern Peoples initiated work that resulted in a Latin-based orthography, developed by Z. Chernjakov, accepted by Narkompros RSFSR in May, 1931. This orthography was in use until 1937, when it was replaced by a Cyrillic orthography, developed by A. G. Endjukovskij. Until this point, the development of the Sámi orthography has followed a path similar to all other non-Slavic languages of the Soviet Union. After WWII, almost all these languages carried on using their newly developed Cyrillic orthographies, with the exception of the deported nationalities (Crimean Tartars, etc.), and of the nationalities with close relative in Finland (Karelians, Vepsians and Sámis).

As for the Sámis, work was initiated in the early 70s to reintroduce the Sámi language in schools, according to school authorities because it was observed that the Sámi children did not master the Russian language properly. It was quickly realized that, contrary to the 1931 orthography, the 1937 Cyrillic orthography did not match the phonemic structure of the Kildin Sámi language, and a new orthography was made, and formally accepted in 1982.

Cyrillic characters in ISO/IEC 10646

Almost all Cyrillic characters of the former Soviet Union are included in 10646. The ones missing are exactly the ones that were not in use in the decades following WWII. Thus, these characters probably were missing from the sources that were used in the preparatory work for the Cyrillic part of 10646 in the first place.

The structure of the characters

CYRILLIC LETTERS SHORT I, EL, EM, WITH DESCENDER, and CYRILLIC LETTER ER WITH TICK

10646 already contains characters with descenders, one of them (CYRILLIC EN WITH DESCENDER) is in use in Kildin Sámi. The descenders cannot be composed, thus they must be included in 10646 as is. The same goes for the CYRILLIC LETTER R WITH TICK. The tick is attached to the basic symbol, its form is unique (there are no other letters composed by that exact diacritic mark), thus no diacritic mark will be able to match the tick of the ER.

CYRILLIC LETTER E WITH DIAERESIS

All WITH DIAERESIS characters can be composed, but in this case it would create a undesirable asymmetry, since all the other Cyrillic (and Latin) characters with diaeresis already have unique non composed positions. If CYRILLIC LETTER E WITH DIAERESIS should be treated as a composed character, the result for Kildin Sámi would be that one of its diaeresis letters would be created directly, the other via composition, a clearly undesirable state of affairs. Of the cyrillic letters denoting vowels, only CYRILLIC LETTER E has no variety with diaeresis 1). As the situation is today, this is a hole in the structure of the Cyrillic subset of 10646.

User community

According to the 1989 Soviet census, (Vestnik Statistiki 1/1989) there are 1888 Kildin Sámi, of which 1.001 has Sámi as their mother tongue.. Kildin Sámi is spoken on the Kola peninsula, in Murmansk Oblast' of North Western Russia.

Sámi is a school subject in primary schools, and articles are occasionally published in Sámi in the local newspaper Lovozerkaja Pravda. As a result of the opening of the borders, the Kildin Sámis are now involved in international Sámi cooperation, among other things in the Sámi Council (active in Russia, Finland, Sweden and Norway, with its main secretariat in Finland).

There is a long scholarly tradition of research on the Sámi (as well as other Uralic) languages, with important research centres including Murmansk, Helsinki, Uppsala, Tromsø, Hamburg, Bloomington, Budapest, to mention a few of them. These institutions regularly publish materials on Kildin Sámi.

The Kildin Sámi literary language is in use in schools, in dictionaries, books and magazines, in international cooperation (the Kildin Sámi are one of the few minorities of Russia that have relatives abroad). Many of the Kildin Sámi books are currently being printed in Norway. The language is also in scientific use in Russia and abroad.

Issues

Importance of 10646 status

To be included in the Basic Multilingual Plane has a value in itself by virtue of serving as a reference point for the letters. Since material in Kildin Sámi is printed in different countries, this will make possible the exchange of manuscripts across the borders, and it will facilitate the printing process. International organisations such as the Sámi Council and the Sámi parliaments will also be able to publish material on Kildin Sámi via their web sites. For the 10646 standard the filling of a hole in the standard will also be important. The goal of the Basic Multilingual Plane is to represent the letters of the written languages currently in use in the world today. When it comes to written languages based on the Cyrillic and Latin alphabets, the coverage is already so good that it takes a small amount of space to make it perfect. 1) Cf. the following table. Letters denoting consonant + vowel sequences have been left out as irrelevant, thus only letters denoting vowels are shown.

vowel	without diaeresis	with diaeresis
A	0410, 0430	04D2, 04D3

E	042D, 044D	
I	0418, 0438	04E4, 04E5
O	041E, 043E	04E6, 04E7
U	0427, 0443	04F0, 04F1
YERU	041E, 044B	04F8, 04F9
SCHWA	04D8, 04D9	04DA, 04DB
BARRED O	04E8, 04E9	04EA, 04EB

The missing diaeresis for E has consequences beyond Kildin Sámi: It makes it harder to use the Cyrillic alphabet in a symmetric way in e.g. dialectology. Assigning a special sound value to the WITH DIAERESIS vowel symbols is problematic when just one of the vowel symbols does not have any WITH DIAERESIS option.

Names and code table

04C5	CYRILLIC CAPITAL LETTER EL WITH DESCENDER	Л	л
04C6	CYRILLIC SMALL LETTER EL WITH DESCENDER		
04C9	CYRILLIC CAPITAL LETTER ER WITH TICK	Р	р
04CA	CYRILLIC SMALL LETTER ER WITH TICK		
04FA	CYRILLIC CAPITAL LETTER E WITH DIAERESIS	Э	э
04FB	CYRILLIC SMALL LETTER E WITH DIAERESIS		
04FC	CYRILLIC CAPITAL LETTER SHORT I WITH DESCENDER	Й	й
04FD	CYRILLIC SMALL LETTER SHORT I WITH DESCENDER		
04FE	CYRILLIC CAPITAL LETTER EM WITH DESCENDER	М	м
04FF	CYRILLIC SMALL LETTER EM WITH DESCENDER		

Téir go dtí innéacs EGT (Go to the EGT index)

Michael Everson, everson@indigo.ie, Dublin, 1997-06-09



Trond Trøsterud
Sjengristisk Institutt
N-7037 Universitetet i Tromsø
+47-77-644239

The Sami Sector at the Russian
Academy of Sciences, Murmansk Branch,
hereby states that the letters repertoire
needed in order to write Kildin Sami
is the following:

Second Testimony 2.

The basic letters are cyrillic letters.

Kildin Sami is used natively only in Russia (with 1889 persons claiming Sami as their nationality in the 1989 Soviet census), but as a part of international Sami cooperation, Kildin Sami books and news items are being printed in Norway and Finland as well. This international aspect makes an international coding standard for Kildin Sami very important indeed.

Rimma Kurneč



KS 14. JUL 1995 13:42

011 HON FAX 47 77044239

14 JUL 96

NO. 0414 NO P. P. 01

47 51295 10580

Аа а Ёё Вв Гг Дд
Ее Ёё Жж Зз Ии Йй
й Кк Лл л Мм м
Нн н н Оо Пп Рр
р Сс Тт Уу Ѡѡ Хх
Цц Чч Шш Щщ ъ Ъ
ь Ъ Ээ Э Юю Яя
Јј Нн , (апостроф)



ISO/IEC JTC1/SC2/WG2

Cyrillic Sámi

As the Norwegian national Body for IT-standardization, NTS is concerned about IT-standardization of the Sámi languages. Due to increased international cooperation among national minorities, the different Sámi languages are now increasingly being used all over the Sámi area. One of the Sámi languages, Kildin Sámi, belongs to the decreasing group of languages of the world that are not fully covered by the repertoire of ISO/IEC10646, as 10 characters of its alphabet are currently missing from this standard.

NTS sees it as parts of its work for the Sámi languages to participate in international efforts to correct this situation. We thus promote the following proposal:

“Proposal to add 10 Cyrillic Sámi characters to ISO/IEC 10646”

Enclosed is also a statement from the head of the Sámi Sector of the Murmansk Branch of the Russian Academy of Science, Ms. Rimma Kuruch, on the Kildin Sámi repertoire.

Yours sincerely

Knut Lindelien

cc.

The Unicode Consortium, P.O.Box 700519, San Jose, CA 95170 - 0519, USA
Forskningssentralen för de inhemska språken, Sörnäs strandväg 25, FIN-00500 Helsingfors
Trond Trosterud, Lingvistisk institutt, Det humanistiske fakultet, N-9037 Universitetet i Tromsø

