

Proposal to encode Latin Capital and Small letter I with Bowl

L2/22-118R

in ISO/IEC 10646 and The Unicode Standard

Nikita Manulov, pacancuvak15@gmail.com

2022-07-23

This proposal is to add the additional Latin letter I with Bowl of Soviet languages

1. Introduction

Ь ь (I with bowl) is an additional letter of the Latin alphabet. It was introduced in 1928 into the reformed Yañalif, and later into other alphabets for Soviet minority languages. The letter was designed specifically to represent the non-front close vowel sounds IPA: [ɨ] and IPA: [ʉ].[1] Thus, this letter corresponds to the letter ⟨I ı⟩ in modern Turkic alphabets

The letter was originally included in the Yañalif, and later also included in the alphabets of the Kurdish, Abazin, Sami, Komi, Tsakhur, Azerbaijani and Bashkir languages, as well as in the draft reform of the Udmurt alphabet. During the project of the Latinization of the Russian language, this letter corresponded to the Cyrillic letter ⟨Й й⟩.

In alphabets that used this letter, the lowercase B was replaced by a small capital ⟨Ь⟩ so that there would be no confusion between ⟨Ь⟩ U+0185 and ⟨b⟩ U+0062.

2. Use of Cyrillic-like Letterforms in Newly Designed Latin Alphabets

The concern was raised that encoding the Latin yeru would open the door to “duplicate the whole Cyrillic alphabet as Latin letters”.

In fact, this is not the case.

In the late 1920s and the 1930s (in some cases even earlier, the Latin alphabet was introduced for many minority languages in the Soviet Union. A considerable part of these were extensions of the Janalif, thus containing the Latin i with bowl to denote the sound of the Cyrillic yeru or related vowels.

While it was “natural” that the designers of the extra letters were influenced by the Cyrillic letters they know, they did copy the exact letterforms (especially doing so for both cases) only in rare instances.

In preparation of a proposal to encode such letters (which will not reach a presentable state before 2011), we until now found no evidence. Alphabets of that era apparently were designed avoiding using of Cyrillic-like letterforms, while containing the Latin I with bowl whether they are derived from Janalif or not.

If examples for use of true Cyrillic letterforms are found which are used only in single alphabets for a small community, in fact it can be considered to treat them like the Zhuang tones three and five, i.e. unifying them with the resembling Cyrillic letters.

However, the Latin I with bowl is a letter used throughout the Soviet Union during the lifetime of Janalif, by

small minorities as well as by large ethnic groups like the Tatars.

Thus, it is appropriate to recognize its use as a Latin letters by giving it a separate code point, like the Latin P and the Cyrillic P have separate code points within their scripts.

3.Examples of use

Figure. 1

New Turkic alphabet (Yañalif)

Aa	Bb	Cc	Çç	Dd	Ee	Əə
Аа	Вв	Сс	Чч	Дд	Ее	Әә
ا	ب	ج	چ	د	ه (je=etep)	ه (هه)
Ff	Gg	Oj oj	Hh	Ii	Jj	Kk
Фф	Гг	Оо	Хх	Ии	Йй	Кк
و	گ	و	خ	ي	ي (اي)	ك
Ll	Mm	Nn	Ŋŋ	Oo	Өө	Pp
Лл	Мм	Нн	Ŋŋ	Оо	Өө	Пп
و	م	ن	ڭ (اولى)	و (اولى)	و (اولى)	پ
Qq	Rr	Ss	Şş	Tt	Uu	Vv
Qq	Рр	Сс	Шш	Тт	Уу	Вв
و	ر	с	ш	т	و (اورون)	و (اد)
Xx	Yy	Zz	žž	Ьь		
Хх	Уу	Зз	Жж	Ьь		
с	و (الوزم)	ز	ж	و (اللى)		

Унифицированный новый тюркский алфавит

Figure. 2

The Latin-based Unified Northern Alphabet

Алфавит Народов Севера
Печатные формы букв

Aa	Bb	Cc	€€	Dd	Зз
Зз	Ee	Əə	Ææ	Ff	Gg
Hh	h̥h̥	Ii	Ьь	Jj	Kk
Ll	Łł	Mm	Nn	Ŋŋ	Oo
Өө	Pp	Qq	Rr	ƦƦ	Ss
Šš	Tt	Uu	Vv	Ww	Xx
Yy	Zz	ƷƷ			

Figure. 3

Kurdish alphabet of 1929

Әлфвaje tәzә kurmançî
НОВЫЙ КУРДСКИЙ АЛФАВИТ

Aa	Bb	Cc	€€	Çç	Dd
А	Б	Ч	сәгм (кожа)	çer (кувшин)	Д
Ee	Əə	Ff	Gg	Qq	Hh
Е	МЯГКОЕ А әм (мы)	Ф	Г	Г МЯГКОЕ тагвәр (кладбище)	ХЫ нәсп (лошадь)
Ii	Ьь	Jj	Kk	ƘƘ	Qq
И	Ы ды (сердце)	Ѕ	К	К кәг (нож)	КЫ qaz (гусь)
Ll	Mm	Nn	Oo	Өө	Pp
Л	М	Н	О	О МЯГКОЕ хоşә (вкусно)	П
Pp	Rr	Ss	Şş	Tt	ƧƧ
Р	Р	С	Ш	Т	Т ТВЕРДОЕ ter (хурджин)
Uu	Yy	Vv	Ww	Xx	
У	МЯГКОЕ духьт (кушак)	В	WƏBŞ (ВЕРЕВКА)	Х	
Zz	ƷƷ				
З	Ж				

Figure. 4

Komi alphabet of 1934



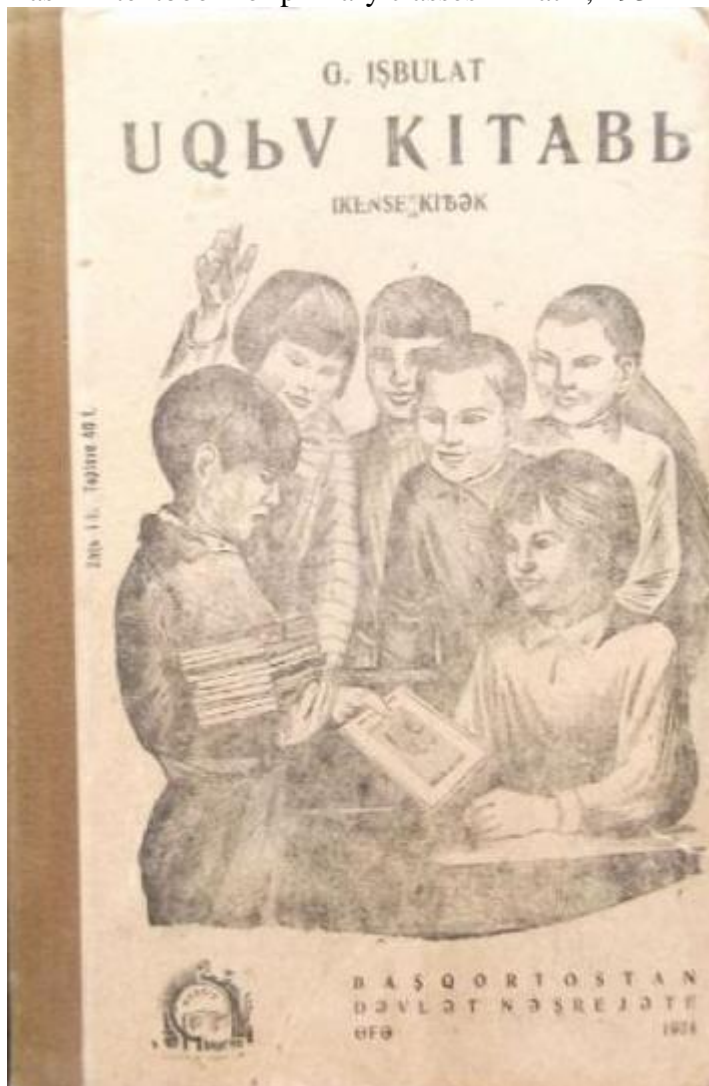
Figure. 5

A Kazakh newspaper Socialist Kazakhstan (Социалды Қазағыстан) in Latin script (1937)



Figure 6.

Bashkir textbook for primary classes in Latin, 1934



[illegible]

— Malajdar jal ittelər kevek, əjdəged qudqalajy
tər, tine Xəlil aqaj, tavəstar verdən janqıranıylar.
— Əjdə Vəlej...
— Əjdə Əxmət.
Ber-ber artı tədelep ytkən havandar artınan qar
kəf təfmalary hudyə vəşlən.

Bol'seviktər.

Sige aləf.
julhəd qola jala
vuşlyqynda vuran ujnayan,
aq haqallıq,
unda qıraqajlyq
tujyn tujlayan,
tarqav tormoş qara
qosaqynda
jata ine ilden
kyp jere,
hələkətkə vayan

Figure 7.

Advertising on Yanalif in the Uzbek magazine “Yangi Yul”, 1933

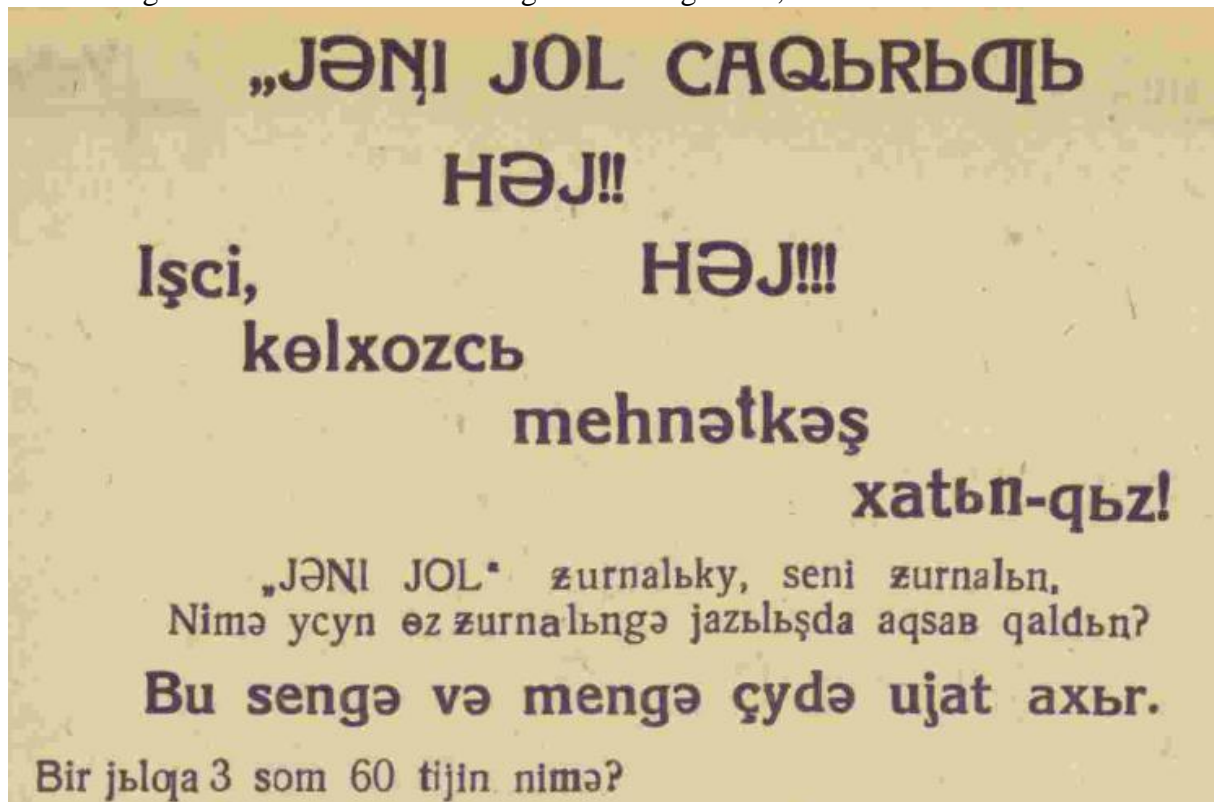


Figure 8.

The frontispiece of of kid's fairies book in kyrgyz language by latin alphabet used in 1930's in Soviet Kyrgyzstan



Figure 9.

5 kopecks of the Tuvan People 's Republic of 1934 with inscriptions in Latin



Figure 10.

Table of correspondence of alphabets in Latin (yanalif) and Cyrillic.

ЯНЫ ЭЛИФБЕ							
№	Янъ elifbæ	Eski elifbæ	Harif-nin adъ	№	Янъ elifbæ	Eski elifbæ	Harif-nin adъ
1	А а	А а	А	20	П п	Р р	Пе
2	Б б	В в	Бе	21	Р р	Р r	Эр
3	В в	V v	Ве	22	С с	S s	Эс
4	Г г	G g	Ге	23	Т т	T t	Те
5	Гъ гъ	Ој ој	Гы	24	У у	U u	У
6	Д д	D d	Де	25	Ф ф	F f	Эф
7	Е е	E e	Е	26	Х х	X x	Ха
8	Ё ё	Jo ø	Ё	27	Ц ц	Ts ts	Це
9	Ж ж	Z z	Же	28	Ч ч	C c	Че
10	З з	Z z	Зе	29	ДЖ дж	Ç ç	Дж
11	И и	I i	И	30	Ш ш	Ş ş	Ша
12	Й й	J j	Й	31	Щ щ	Şc şc	Ща
13	К к	K k	Ка	32	Ъ ъ	ajъrъcъ ve qattъ- lъq işa- retî	Айыры- лжы ве къатты- лыкъ иша- ретî
14	Къ къ	Q q	Кы				
15	Л л	L l	Эл	33	Ы ы	Ь ь	Ы
16	М м	M m	Эм	34	Ь ь	inçertyv işareti	инджер- тювиша- ретî
17	Н н	N n	Эн	35	Э э	E e	Э
18	Нъ нъ	N n	Энг	36	Ю ю	Ju jy-y	Ю
19	О о	O o	О	37	Я я	Ja ja	Я

Figure 11.

A Azerbaijani poem written by Mehdi Huseyn-zadeh in April 1944 in Latin

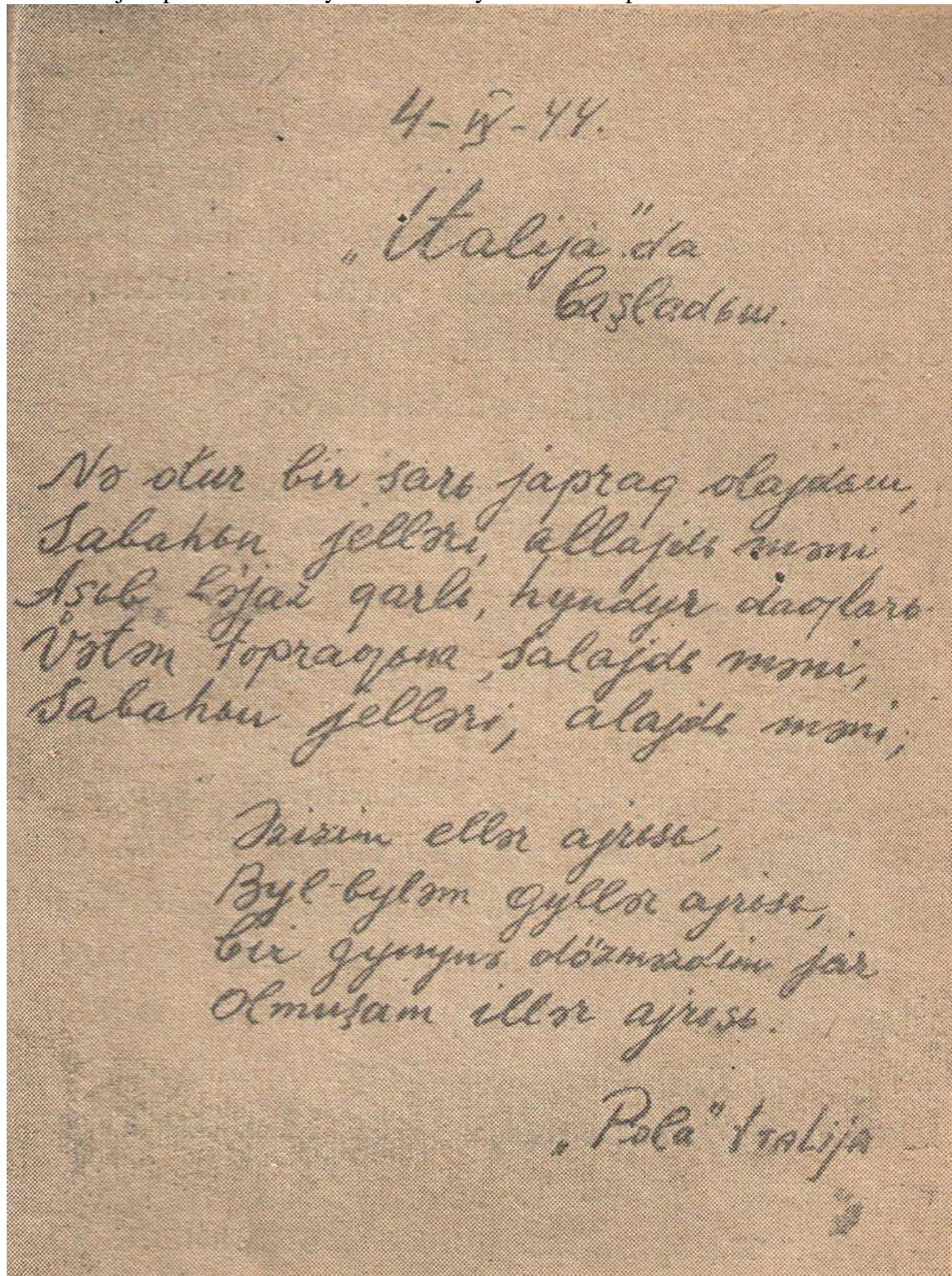


Figure 12.

Altaic Romanized Alphabet (1929-1938)

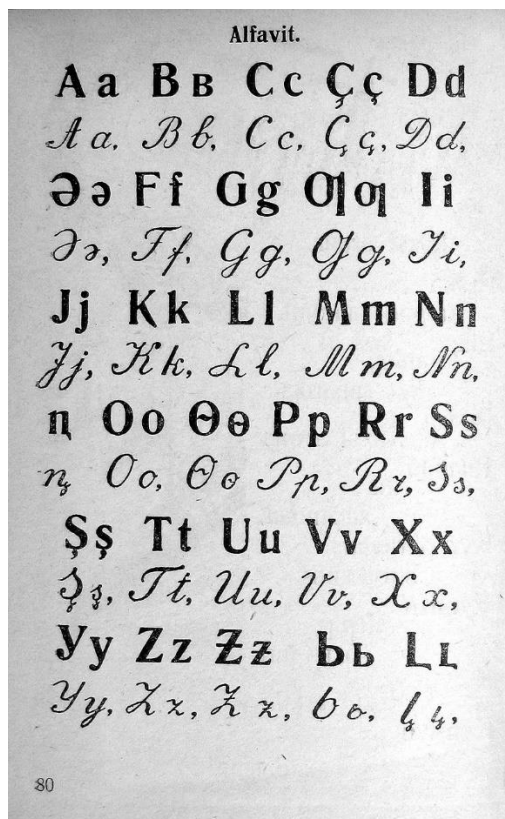
Çaңь Altaj biciktös „jaңalif“ Новый алтайско-Тюркский алфавит						
A a А	B b Б	C c Ч	Ç ç дь	D d Д	E e Э	F f Ф
G g Г	I i И	J j Й	K k К	L l Л	M m М	N n Н
Ŋ ŋ Ң	O o О	Ө ө Ө	P p П	R r Р	S s С	Ş ş Ш
T t Т	U u У	Y y Ү	Z z З	Ž ž Ж	Ь ь Ы	X x Х
V v В	ja Я	ju Ю	je Е	ts Ц	şç Щ	

Figure 13.

The Kumyk alphabet on a Latin basis from the primer of 1935

Qumuq alifba.					
A a А а	B b Б б	C c С с	Ç ç Ç ç	D d Д д	E e Е е
F f Ф ф	G g Г г	Q q Қ қ	H h Һ һ	I i И и	J j Ј ј
K k К к	L l Л л	M m М м	N n Н н	N n Ң ң	O o О о
Ө ө Ө ө	P p П п	Q q Қ қ	R r Р р	S s С с	Ş ş Ш ш
S s С с	T t Т т	U u У у	V v В в	X x Х х	Y y Ү ү
Z z З з	Z z З з	b b Б б	ba'li Ба'ли		

Figure 14.
Scan form Khakas book 'Bukvar' (1934)



4. Characters to be added

Block: Latin Extended-D																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+A7Cx	?	?	?	?	?	?	?	?	?	?	?				?	?

Character Code point Name

Ї U+A7CE LATIN CAPITAL LETTER I WITH BOWL

→ Ї 042B cyrillic capital letter yeru

→ Ь 042C cyrillic capital letter soft sign

→ Ъ 0184 latin capital letter tone six

ѣ U+A7CF LATIN SMALL LETTER I WITH BOWL

→ і 0131 latin small letter dotless i

5. UnicodeData.txt

A7CE; LATIN CAPITAL LETTER I WITH BOWL; Lu;0; L;;;; N;;; A7CD;
A7CF; LATIN SMALL LETTER I WITH BOWL; Ll;0; L;;;; N;;; A7CC;; A7CC

6. References

- 1.М. З. Закиев. Тюрко-татарское письмо. История, состояние, перспективы. Москва, «Инсан», 2005
2. Proposal to encode two Latin letters for Jaʼalif (WG2 N3916) Karl Pentzlin, 2010
3. М. Temirxanov. Alifba. — Mahac-Qala, 1935.
4. Samrin K.K. Bukvar, 1934
5. <https://bashkort-tele.livejournal.com/79147.html#cutid1>
6. no. Yangi Yol magazine, 1933
- 7.Baldar Çomoqtoru. Aјtuucu: Toqoloq Moldo Abdьraqmanov. 1939
- 8.no. Совет Къадынлыгы. 1938, № 7

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to encode Latin Capital and Small letter I with Bowl		
2. Requester's name:	Nikita Manulov		
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution		
4. Submission date:	2022-05-20. revised 2022-07-23		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:	Yes		
(or) More information will be provided later:			

B. Technical – General

1. Choose one of the following:			
a. This proposal is for a new script (set of characters):			No
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:			Yes
Name of the existing block:	Latin Extended-D		
2. Number of characters in proposal:			2
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	B.1-Specialized (small collection)	x	B.2-Specialized (large collection)
C-Major extinct	D-Attested extinct		E-Minor extinct
F-Archaic Hieroglyphic or Ideographic			G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?			
			Yes
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?			
			Yes
b. Are the character shapes attached in a legible form suitable for review?			
			Yes
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?			
			Nikita Manulov
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:			
https://drive.google.com/file/d/1ICMjSxDJUwDJM3pigDwr2V8r6aUJug2-/view?usp=sharing			
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			Yes
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?			Yes
7. Special encoding issues:			
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?			No

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N3152-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	YES
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	NO
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	See proposal
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	Common
Reference:	
5. Are the proposed characters in current use by the user community?	historical
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	Yes
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	No
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	No
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	No
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	

