Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type:	**Working Group Document**
Title:	**Proposal to encode one historical Mongolian letter for Buryat Mongolian**
Source:	**Andrew West, Amgalan Zhamsoev (Institute of Mongolian, Buddhist and Tibetan studies SB RAS, Ulan-Ude, Buryatia, Russia), Viacheslav Zaytsev (Institute of Oriental Manuscripts RAS, Saint Petersburg, Russia)**
Status:	**Individual Contribution**
Action:	**For consideration by JTC1/SC2/WG2 and UTC**
Date:	**2017-01-13**

1. Introduction

This is a proposal to encode an additional Mongolian letter that was historically used for writing texts in Traditional Mongolian script by the Buryats of the Barguzin Steppe Duma in Russia. This letter is no longer used, but is required for encoding because scholars studying documents written in Traditional Mongolian script from Barguzin wish to be able to represent text using this letter in Unicode Mongolian.

The letter is a form of Mongolian Letter Cha ᡁ (U+1834) with two dots placed to the right of the stem (ᡂ). The use of two dots to distinguish ᡂ from U+1834 ᡁ is influenced by the use of two dots in Mongolian Letter Sha ᠶ (U+1831) to distinguish it from Mongolian Letter Sa ᠰ (U+1830).

This dotted form of the letter cha is used to indicate that it is pronounced as š rather than č, reflecting Buryat pronunciation practice which differs significantly from Classical Mongolian. Rinčen 1965 discusses the Buryat pronunciation of č and š (see Fig. 1), which we translate below:[1]

> 10. In the Mongolian script Buryats and some of the southern Mongols living in part of the former Jirem League do not distinguish in pronunciation [the characters] č and š of Mongolian script. When they spell out the character č, the Buryats call it *eberei ša* "front š", and southerners (Gorlos and others) [call it] *bosō ša* "standing š"; and the character š, the Buryats [call it] *arae ša* "back š", and southerners [call it]

[1] This note by Rinčen is given for the following part from the Buryat Genealogy document he studies: *Qayiča qar-a bayatur kümün aysan bui* "Хайса Хара был богатырём" [Khaysa Khara was a bogatyr]. It explains the reason why the name spelled in the document as Qayiča qar-a was read as Qayisa qar-a "Хайса Хара" [Khaysa Khara] by him.

χeptē ša “reclining š”, [due to] the appearance of these characters, as č has a hook in front, which sticks out or stands up, and the character š faces back at a reclined angle. This is why, in blockprints of the Buryats, Gorlos and Khorchin, so often the closing word *tegüsbe* “end” at the end of the book is found in the form of *tegüčbe*. Therefore, I think that Qayisa qar-a [actually Qayiča qar-a] of our Genealogy [the document Rinčen studies in this paper] should be read as Khaysa Khara, following the Buryat pronunciation, not the spelling [i.e. Qayiča qar-a]. Perhaps that is a truncated form of the Mongol name Khaysan Khara (Qayisang qar-a).

Tsydendambaev 1972 also discusses this linguistic phenomenon (See Fig. 2), which we translate below:

> Although the intention to somehow distinguish the letters g and k, which are homonymic in their graphic forms in Classical Mongolian language ( ), has a certain cause, in spoken language they developed phonetic distinctions. In the “History of migration to Barguzin…”, by placing dots on the right of character *ṣādhē* ( ) it is shown that the Mongolian letter č in Buryat language is pronounced with the sound š.

Tsydendambaev 1972 gives two examples of Buryat texts where the dotted letter cha is used to indicate the Buryat pronunciation of š (see Fig. 3), which we translate below:

> 4. In the documents of Barguzin Steppe Duma, the character *ṣādhē* with dots at the right ( ) was made use of for representing the letter š: (fond 7, delo 76, folio 191) instead of *egün-eče doγuγši* ‘somewhat lower than this place’; (fond 7, delo 321, folio 103) instead of *egün-ü doruγši-du* ‘below this’; and some other [examples].

Digital images of the manuscripts showing these two examples, published here for the first time by kind permission of the State Archive of the Republic of Buryatia, are given as Figs. 6 and 7 below.

Tsydendambaev 1972 gives several additional examples from a manuscript of the Barguzin Buryat chronicle “History of the migration [that happened] formerly of the Barguzin Buryats from the north of the Ocean [Lake Baikal] to Barguzin under the command of Ondrey Shibsheev in the year 1740” (*Barγuǰin-u buriyad-nar-un erete [=erte] urida dalai-yin qoyin-a-ača 1740 on barγuǰin-du onderui sibsig-e-ün türtei [=türütei] negüǰü iregsen teüke domoγ bui*; known in Russian translation as “История перекочёвки в Баргузин в 1740 году баргузинских бурят с севера Байкала под предводительством Ондрея Шибшеева”), written by Tsedebzhab Sakharov at the end of the 19th century (see Fig. 4).[2] We translate below:

> Orthographical and grammatical variant readings, that are to be found in particular [Buryat] historical works, are of local and inconsequent kind. For example, an anonymous author [identified as Ts. Sakharov later] of one of the Barguzin

---

[2] This manuscript is held by the Institute of Mongolian, Tibetan and Buddhist studies SB RAS (M I collection, inv. No. 21, old No. 373). See Tsyrempilov 2004, p. 3, No. 1.

manuscripts (“History of migration to Barguzin…”), in his attempt to show Buryat pronunciation graphically, designates hushing sound š with character *ṣādhē* with dots on the right ():  *činagši* instead of regular  *činaɣši* ‘further’,  *ošiɣsan* instead of  *očiɣsan* ‘departed’,  *bišin* instead of  *bičin* ‘monkey’ and so on [Ms. inv. No. M-I-373, f. 2, 7 verso and others]. But along with this in the same manuscript the same consonant š is designated, as it is usually done in Mongolian graphics, with the character *šīn* with dots on the right (): for example,  *šibše-e* (Buryat pronunciation of the name of the Barguzin taishas’ ancestor Čibčiɣai),  *tayiša-a* ‘taisha’, *qašilta* ‘oppression’ [*Ibid*., f. 2 and others].

The use of the dotted cha letter is not restricted to the writings of one or two authors, quoted from above, but can be found in other Buryat manuscripts. Fig. 5 shows the occurrence of this letter in the Mongolian transcription of a Tibetan place name in a manuscript map from Buryatia.

The examples of Buryat Mongolian (BM) words written with the proposed letter () that are given in Figs. 3–5 are summarised below, with their equivalent spellings in Classical Mongolian (CM):[3]

- Fig. 3: BM:  (*doɣugṣi*) = CM:  (*doɣuɣsi*)
- Fig. 3: BM: (*dorugṣid*) = CM: (*doruɣsi-du*)
- Fig. 4: BM:  (*činagṣi*) = CM: (*činagsi*)
- Fig. 4: BM:  (*oṣiɣsan*) = CM: (*očiɣsan*)
- Fig. 4: BM:  (*biṣin*) = CM: (*bičin*)
- Fig. 5: BM:  (*naɣṣiu*) = Tibetan ནག་ཆུ (*nag chu*)

[3] In the transcription of Buryat Mongolian, we transcribe the proposed character as ṣ̈ in order to distinguish it from š which transcribes the letter .

2. Distinct Letter or Variant Form?

It may be questioned whether this letter should be encoded as a new character or whether it should be defined as a variant form of U+1834 ᠴ (i.e. 1834 + FVS1). We consider that it would be most appropriate to encode it as a new character. Firstly, the character is distinguished from U+1834 by two dots on the right of the stem, and there are no other cases where two dots on the right are used to indicate variation; and on the contrary in the case of U+1831 ᠱ the two dots on the right indicate that it is a separate character from U+1830 ᠰ. The proposed character ᠴ᠋ can be considered to be a new letter combining the body of U+1834 ᠴ and the dots of U+1831 ᠱ, and not a variant of either U+1834 or U+1831. Secondly, the proposed character explicitly has a different pronunciation to U+1834, and none of the existing Mongolian variation sequences indicate a pronunciation difference. We consider that the pronunciation difference between ᠴ and ᠴ᠋ indicates that the latter is a distinct letter rather than a variant. The example of ᠴᠢᠨᠠᠭᠱᠢ (*činagṣi*) given in Fig. 4 is a good case in point, where the letters ᠴ and ᠴ᠋ are used contrastively, the former pronounced č, and the latter pronounced š, which indicates that they cannot be treated as simple glyph variants.

3. Unicode Properties

We suggest encoding the proposed character at U+1878, which is the first reserved space after the basic Mongolian letters (and before the *ali gali* letters). The suggested character name is "MONGOLIAN LETTER CHA WITH TWO DOTS" which describes its shape.

The proposed letter has the same positional shaping behaviour as U+1834, although we have found no examples of the initial or final forms.

Proposed Character and Positional Forms

1878	MONGOLIAN LETTER CHA WITH TWO DOTS	
	Initial form (not attested)	
	Medial form	
	Final form (not attested)	

UCD Properties

1878;MONGOLIAN LETTER CHA WITH TWO DOTS;Lo;0;L;;;;;N;;;;;

Line break: AL

Joining type: D (Dual Joining)

Script: Mongolian

4. *Figures*

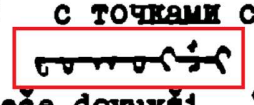
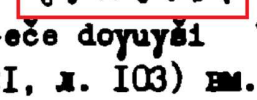
Fig. 1: Rinčen 1965 p. 212 footnote 10

[10] В монгольском письме буряты и некоторая часть южных монголов, населяющих часть бывшего Джиримского сейма не различают в произношении *č* и *š* монгольского письма и называют при спеллинге знак *č* буряты — *eberei ša* «передним *š*», а южане — горлосцы и др. *bosō ša* «стоячим *š*», знак *š* буряты *arae ša* «задним *š*», южане — *χeptē ša* «лежачим *š*», по внешнему виду этих знаков, так как *č* имеет крючок впереди, который торчит или стоит, а знак *š* лежачим углом обращен назад. В бурятских, горлоских и хорчинских ксилографах поэтому очень часто в конце книг заключительное слово *tegüsbe* «конец» встречается в форме *tegüčbe*. Поэтому я думаю, что *Qayisa qar-a* нашей родословной следует читать как Хайса Хара, следуя бурятскому произношению, а не написанию. Возможно, что это усеченная форма монгольского имени Хайсан Хара — *Qayisang qar-a*.

Fig. 2: Tsydendambaev 1972 p. 304

диакритический знак. Правда, стремление как-то разграничить буквы g и k, омонимичные по своему изображению в классическом монгольском языке (ᠭ · ᠭ · ᠺ), имеет известное основание: в живой речи они приобрели фонематические различия. В "Истории перекочевки в Баргузин..." путем постановки точек справа от знака ṣādhē () показано, что монгольская буква č в бурятском языке произносится как звук š.

Fig. 3: Tsydendambaev 1972 p. 556

4. В бумагах Баргузинской степной думы употреблен знак ṣādhē с точками справа () для изображения буквы š:  (ф. 7, д. 76, л. 191) вм.  egün-eče doγuγši 'пониже этого места'; (ф. 7, д. 321, л. 103) вм.  egün-ü doruγši-du 'ниже сего' и некоторые другие.

See Figs. 6 and 7 for the manuscript sources for these examples

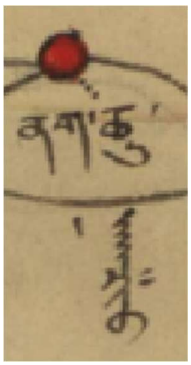
Fig. 4: Tsydendambaev 1972 pp. 297–298

Некоторые отличия в области правописания и употребления грамматических средств, встречающиеся в отдельных исторических сочинениях, носят частный и непоследовательный характер. Так, например, анонимный автор одной из баргузинских рукописей - "Истории перекочевки в Баргузин...", стремясь изобразить бурятское произношение графически, обозначает шипящий согласный š знаком ṣādhē с точками справа (ṣ): činagši вм. обычного činayši 'дальше', ošiysan вм. očiysan 'отправился', bišin вм. bičin 'обезьяна' и т. п.[I] Но вместе с тем в этой хронике тот же согласный š обозначается, как и обычно в монгольской графике, знаком šīn с точками справа (ш): например, šibše-e - бурятское произношение имени предка баргузинских тайшей Чивчегуя, tayiša-a 'тайша', qašilta 'притеснение'[2]. В "Адагат-

[I] РО БФ СО АН СССР, инв. № М-I-373, л. 2, 7 об. и др.

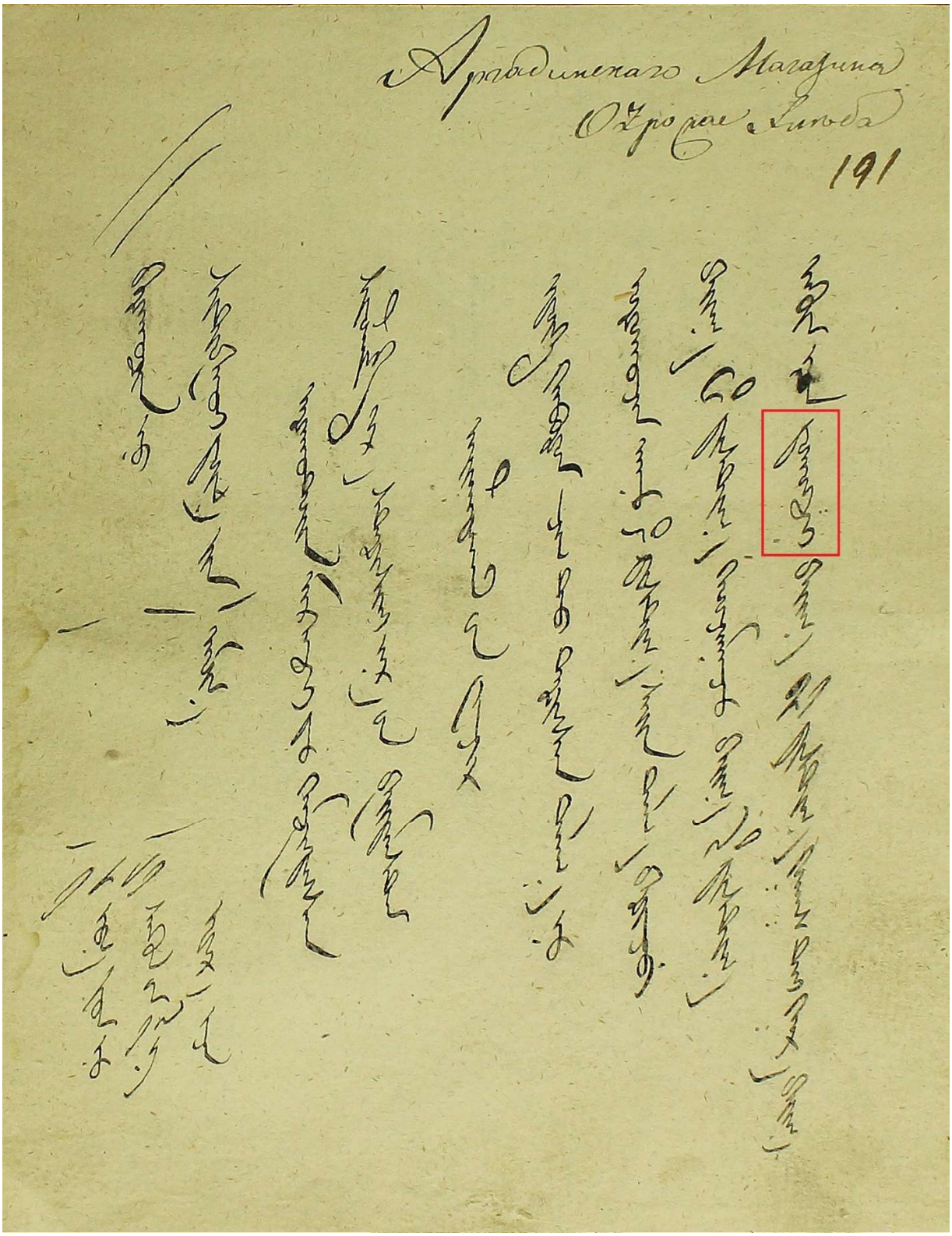
[2] Там же, л. 2 и др.

Fig. 5: Detail of a Mongolian manuscript pilgrimage map from Buryatia (courtesy Jargal Badagarov)



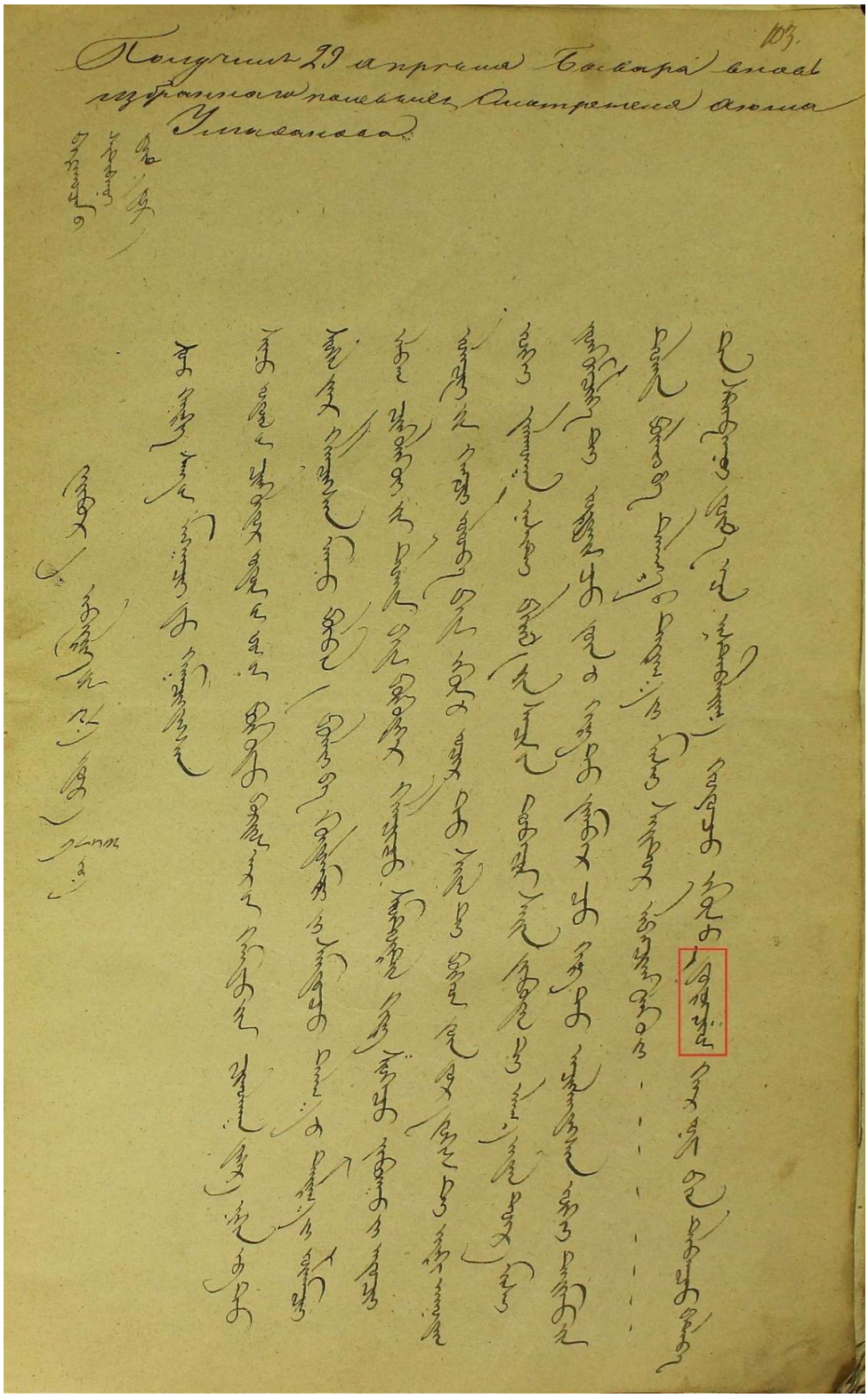
Tibetan: ནག་ཆུ (*nag chu*); Buryat Mongolian: ᠨᠠᠶᠢᠰᠢᠤ (*nayšiu*)

Fig. 6: State Archive of the Republic of Buryatia. Fond № 7 [Barguzin Steppe Duma], file (delo) № 76, folio № 191



Buryat Mongolian: ᠳᠣᠭᠤᠭᠰᠢ (*doγugšï*)

Fig. 7: State Archive of the Republic of Buryatia. Fond № 7 [Barguzin Steppe Duma], file (delo) № 321, folio № 103



Buryat Mongolian: ᠳᠣᠷᠤᠭᠱᠢᠳ (*dorugšid*)

5. Bibliography

Rinčen B. (Ринчен), “Об одной хори-бурятской родословной” [About one Khori-Buryat Genealogy]; *Acta Orientalia Academiae Scientiarum Hungaricae*, T. 18, Fasc. 1/2 (1965): 205–225.

Tsydendambaev, Ts. B. (Цыдендамбаев Ц. Б.), *Бурятские исторические хроники и родословные. Историко-лингвистическое исследование* [Buryat historical chronicles and genealogies. Historic and linguistic study]. Ulan-Ude: Бурятское книжное издательство [Buryat book publishing house], 1972.

Tsyrempilov, N. (Цыремпилов Н. В.) (comp.). *Annotated Catalogue of the collection of Mongolian manuscripts and xylographs M I of the Institute of Mongolian, Tibetan and Buddhist studies of Siberian Branch of Russian Academy of Sciences* = ロシア科学アカデミーシベリア支部モンゴル学・チベット学・仏教学研究所モンゴル語写本・版本 MI コレクション注釈付目録. Sendai: Center for Northeast Studies, Tohoku University, 2004. (CNEAS Monograph Series No. 17)

6. *Proposal Summary Form*

SO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[4]
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 one historical Mongolian letter for Buryat Mongolian*
2. Requester's name: *Andrew West, Amgalan Zhamsoev, Viacheslav Zaytsev*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2017-01-13*
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: *Mongolian*
2. Number of characters in proposal: *1*
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. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   Andrew West
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Andrew West
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 Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[4] Form number: N4102-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, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *NO*
   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.)? *YES*
   If YES, with whom? *public-i18n-mongolian@w3.org*
   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? *NO*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare) *rare*
   Reference:
5. Are the proposed characters in current use by the user community? *NO*
   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)? *YES*
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, or could be confused with, an existing character? *NO*
    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 characters? *NO*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: