

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 add six Syriac letters for Sogdian and Persian to the UCS  
**Source:** Nicholas Sims-Williams and Michael Everson  
**Status:** Expert Contribution  
**Date:** 2002-03-30

## A. Administrative

### 1. Title

Proposal to add six Syriac letters for Sogdian and Persian to the UCS.

### 2. Requester's name

Nicholas Sims-Williams and Michael Everson.

### 3. Requester type (Member body/Liaison/Individual contribution)

Expert contribution.

### 4. Submission date

2002-03-30

### 5. Requester's reference (if applicable)

### 6. Choose one of the following:

#### 6a. This is a complete proposal

Yes.

#### 6b. More information will be provided later

No.

## B. Technical -- General

### 1. Choose one of the following:

#### 1a. This proposal is for a new script (set of characters)

No.

#### Proposed name of script

#### 1b. The proposal is for addition of character(s) to an existing block

Yes.

#### 1b. Name of the existing block

Syriac (proposed code positions: U+072D, U+072E, U+072F, U+074D, U+074E, U+074F)

### 2. Number of characters in proposal

6.

### 3. Proposed category (see section II, Character Categories)

Category A.

#### 4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 1.

#### 4b. Is a rationale provided for the choice?

Yes.

#### 4c. If YES, reference

Base characters with no diacritics.

#### 5a. Is a repertoire including character names provided?

Yes.

#### 5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

#### 5c. Are the character shapes attached in a legible form suitable for review?

Yes.

**6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?**

Michael Everson, Everson Typography. TrueType.

**6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:**

Michael Everson, Everson Typography. Fontographer.

**7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?**

Yes (see below).

**7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?**

Yes.

**8. 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.

**9. 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/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.**

Each of these characters sorts as a separate letter immediately following the original letters from which they were adapted. Hence: BETH, BHETH, GAMAL, GHAMAL, DALATH, DHALATH, ZAIN, ZHAIN, KAPH, KHAPH, PE, FE.

## **C. Technical -- Justification**

**1. Has this proposal for addition of character(s) been submitted before? If YES, explain.**

No.

**2a. 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.

**2b. If YES, with whom?**

Nicholas Sims-Williams is a leading specialist in Iranian languages.

**2c. If YES, available relevant documents**

The proposal is an elaboration of what was agreed at the 2nd Meeting of the Iranianist Unicode group, Frankfurt, 22-23 November 2001, the report of which is an official Unicode document (L2/02-009)

**3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?**

Indo-Iranian, Indo-European, and Turkic philologists and theologians.

**4a. The context of use for the proposed characters (type of use; common or rare)**

Supplementary Syriac letters used to write Sogdian, Persian, and Central Asian Turkish.

**4b. Reference**

See below.

**5a. Are the proposed characters in current use by the user community?**

Yes.

**5b. If YES, where?**

By Indo-Iranian, Indo-European, and Turkic philologists and theologians.

**6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?**

Yes.

**6b. If YES, is a rationale provided?**

Keeping them with other Syriac characters.

**6c. If YES, reference**

**7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?**

No.

**8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?**

No.

**8b. If YES, is a rationale for its inclusion provided?**

**8c. If YES, reference**

**9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?**

No.

**9b. If YES, is a rationale for its inclusion provided?**

**9c. If YES, reference**

**10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?**

No.

**10b. If YES, is a rationale for its inclusion provided?**

**10c. If YES, reference**

**11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?**

No.

**11b. If YES, is a rationale for such use provided?**

**11c. If YES, reference**

**12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?**

No.

**12b. If YES, reference**

**13a. Does the proposal contain characters with any special properties such as control function or similar semantics?**

No.

**13b. If YES, describe in detail (include attachment if necessary)**

**14a. Does the proposal contain any Ideographic compatibility character(s)?**

No.

**14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?**

**14c. If YES, reference**

## D. Proposal

Between about the 8th and 14th centuries CE, the Syriac script was widely used by Christian communities in Central Asia and the Far East to write languages other than Syriac, which they adapted for that purpose by the addition of several additional letters.













The most substantial surviving body of material written in this extended Syriac alphabet consists of manuscripts in Sogdian, a language which belongs to the Iranian branch of the Indo-European family. Sogdian was originally the language of the area around Samarkand (Uzbekistan), but all known Sogdian manuscripts in Syriac script come from much farther east, from sites in the Turfan oasis in Xinjiang (Western China). These manuscripts are now in Berlin, in the keeping of the Berlin-Brandenburgische Akademie der Wissenschaften, and almost all of them can be seen on its web-site: <<http://www.bbaw.de/vh/turfan/index.html>> (select “Digitales Turfan-Archiv” and then “Christlich-soghdische Texte in nestorianischer Schrift”). The most recent survey of the Christian Sogdian literature is by N. Sims-Williams in A. Cadonna (ed.), *Turfan and Tunhuang: the texts*, Florence 1992, pp. 43–61. The writing system has been described by N. Sims-Williams in R. Schmitt (ed.), *Compendium Linguarum Iranicarum*, Wiesbaden 1989, p. 178, with table on p. 176, and (very briefly) by P. Oktor Skjærvø in P. T. Daniels and W. Bright (eds.), *The World's Writing Systems*, Oxford–New York 1996, 533–4.

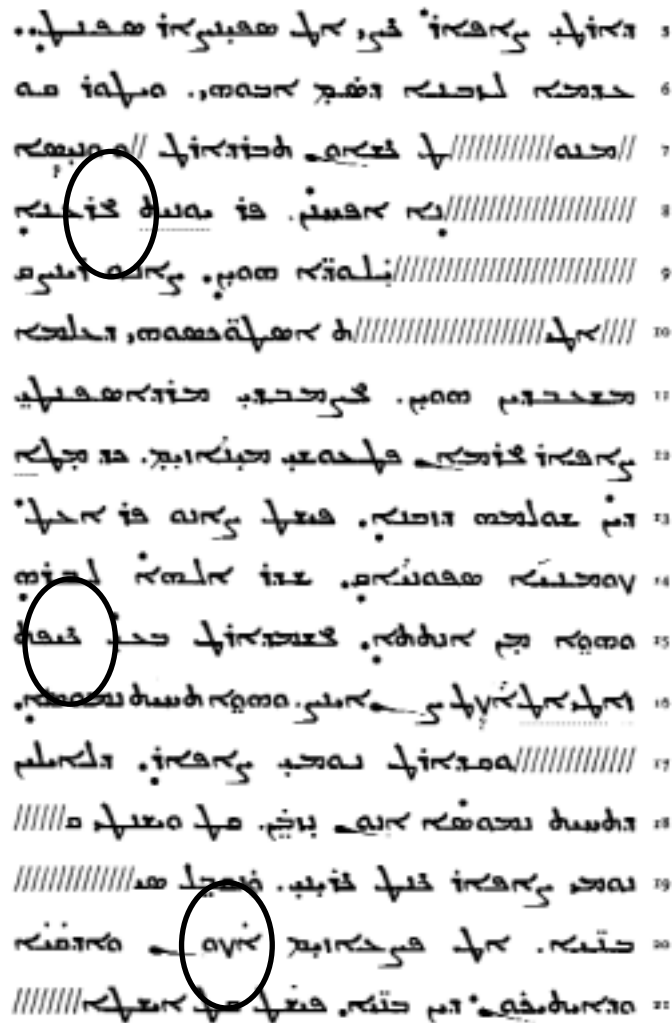
In addition to almost the full repertoire of Syriac letters, vowel points and diacritic marks, the Sogdian manuscripts in Syriac script make use of three additional letters: an adapted ZAIN to represent the palatal [ʒ], an adapted KAPH to represent the fricative [x], an adapted PE to represent the fricative [f]. By analogy, these could be named SYRIAC LETTER SOGDIAN ZHAIN, SYRIAC LETTER SOGDIAN KHAPH, and SYRIAC LETTER SOGDIAN FE.

Of these additional letters, at least KHAPH remained in use long after Sogdian ceased to be written (in about the 10th century), e.g. on the Christian tombstones from Kyrgyzstan and elsewhere inscribed in Syriac mixed with Turkish words and names: many examples in W. Klein, *Das nestorianische Christentum an den Handelswegen durch Kyrgyzstan bis zum 14. Jh.*, Turnhout 2000. All three characters have a typographic tradition dating back to F. W. K. Müller's publications at the beginning of last century, which shows forms designed to harmonize with two different Syriac type-faces (East Syriac and Estrangela).

The Berlin collection also includes a few Persian manuscripts in Syriac script, including fragments of a Psalm-book and of a pharmacological text. These make use of the same extended Syriac script as Sogdian, with the addition of three further additional letters: an adapted BETH to represent the fricative [v], a DALATH with two subscript dots (instead of one) to represent the fricative [ð], and an adapted GAMAL to represent the fricative [ɣ]. These three letters could be named SYRIAC LETTER PERSIAN BHETH, SYRIAC LETTER PERSIAN DHALATH, and SYRIAC LETTER PERSIAN GHAMAL. (The letter GHAMAL is also known to have been used, probably as a result of Persian influence, in a few late Sogdian manuscripts, but all of these seem to have been lost or destroyed in or after the Second World War. In general, Christian Sogdian manuscripts use the Syriac letter E to represent the fricative [ɣ].)

The joining characteristics of all these six letters are the same as those of the Syriac letters on which they are based. The shapes of the letters can be described as follows:

-  SYRIAC LETTER PERSIAN BHETH: same as  BETH with an oblique stroke added at the top left.
-  SYRIAC LETTER PERSIAN GHAMAL: same as  GAMAL with a hook added at the top left.
-  SYRIAC LETTER PERSIAN DHALATH: same as  DALATH with two points below instead of one.
-  SYRIAC LETTER SOGDIAN ZHAIN: based on the  "I"-shaped ZAIN, converted into a "V"-shape.
-  SYRIAC LETTER SOGDIAN KHAPH: same as  KAPH with a hook added at the top left.
-  SYRIAC LETTER SOGDIAN FE: same as  PE with an extension at the top left.



**Figure 1.** Christian Sogdian manuscript (Epistle to the Galatians, chapter 3), showing the letters FE, KHAPH, and ZHAIN, typeset in an Estrangela type-face: from F. W. K. Müller, *Soghdische Texte I* (Abhandlungen der Königlich Preussischen Akademie der Wissenschaften, 1912, No. 2), Berlin 1913, p. 84.

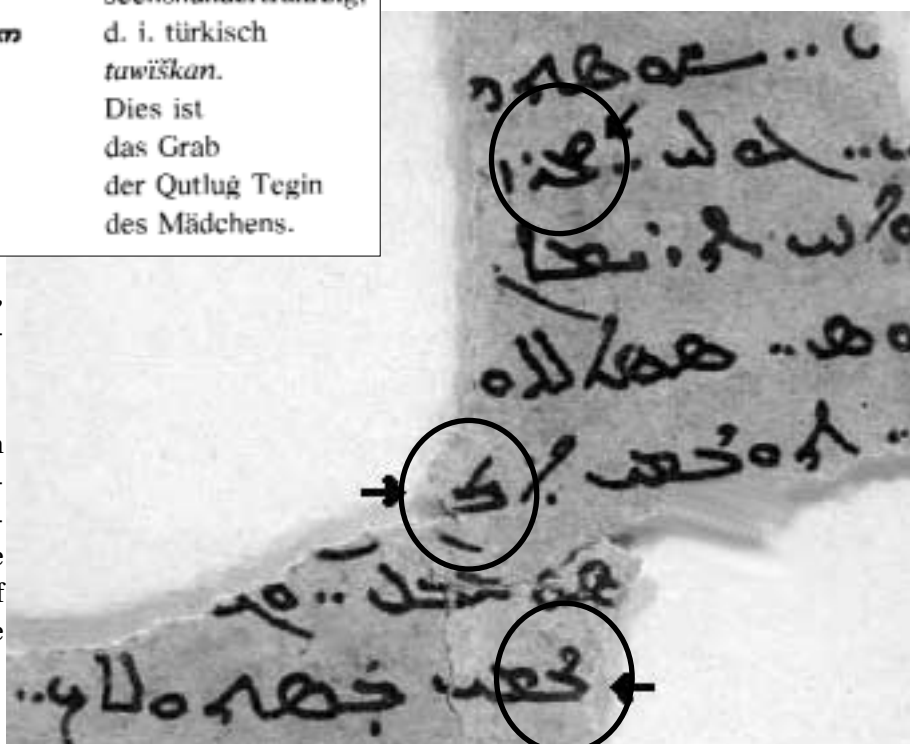


**Figure 2.** Christian tombstone from Kyrgyzstan (1338/9 C.E.), showing the continuing use of the letter KHAPH amongst Turkish Christians in the 14th century: from W. Klein, *Das nestorianische Christentum an den Handelswegen durch Kyrgyzstan bis zum 14. Jh.*, Turnhout 2000, Abb. 64.

/1/ ܘܘܠܐ ܕܘܠܐ	Im Jahr eintausend-
/2/ ܘܘܠܐ ܕܘܠܐ	sechshundertfünfzig,
/3/  sic  ܘܘܠܐ ܕܘܠܐ	d. i. türkisch
/4/ ܘܘܠܐ ܕܘܠܐ	tawişkan.
/5/ ܘܘܠܐ ܕܘܠܐ	Dies ist
/6/ ܘܘܠܐ ܕܘܠܐ	das Grab
/7/ ܘܘܠܐ ܕܘܠܐ ←	der Qutluğ Tegin
/8/ ܘܘܠܐ ܕܘܠܐ	des Mädchens.

**Figure 3** (above). The same, typeset in an Estrangela typeface: *ibid.*, p. 173.

**Figure 4** (right). Christian Persian pharmacological manuscript, showing (top to bottom, indicated by arrows) the letters FE, BHETH, KHAPH: part of <n175rectototal.jpg> on the BBAW web-site.



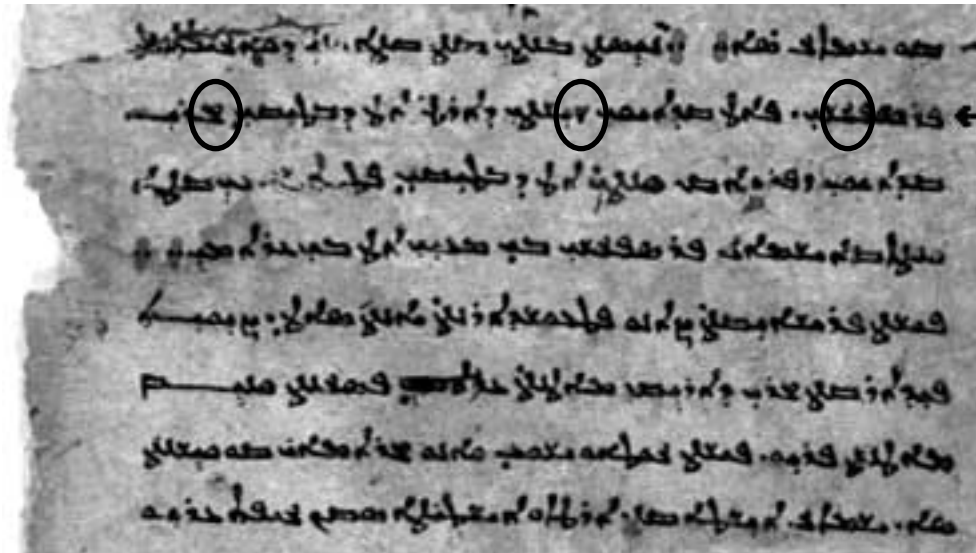


Figure 5. Christian Sogdian manuscript (St Luke's gospel, chapter 16), showing the letters FE, ZHAIN, and KHAPH: part of <n153versototal.jpg> on the BBAW web-site.

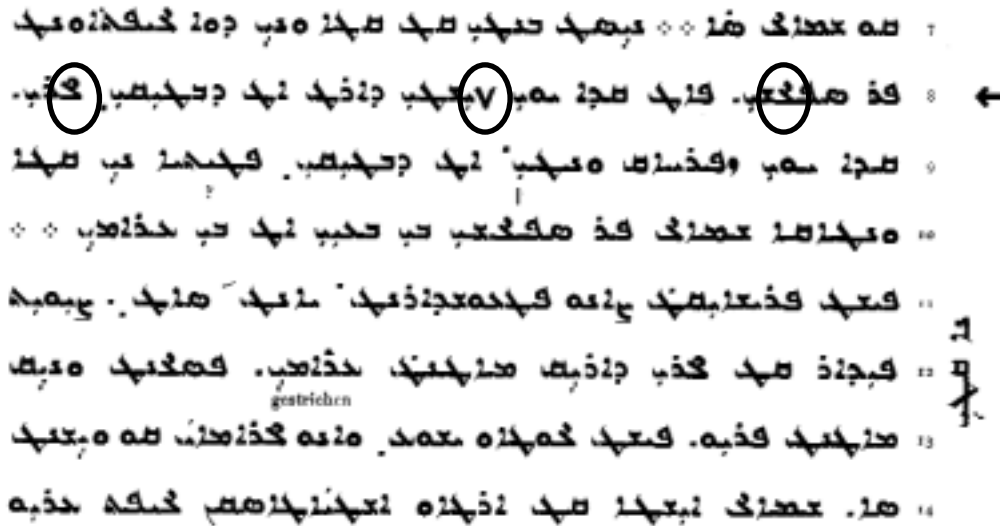


Figure 6. The same, typeset in an East Syriac type-face: from F. W. K. Müller, *Soghdische Texte I* (Abhandlungen der Königlichen Preussischen Akademie der Wissenschaften, 1912, No. 2), Berlin 1913, p. 44.

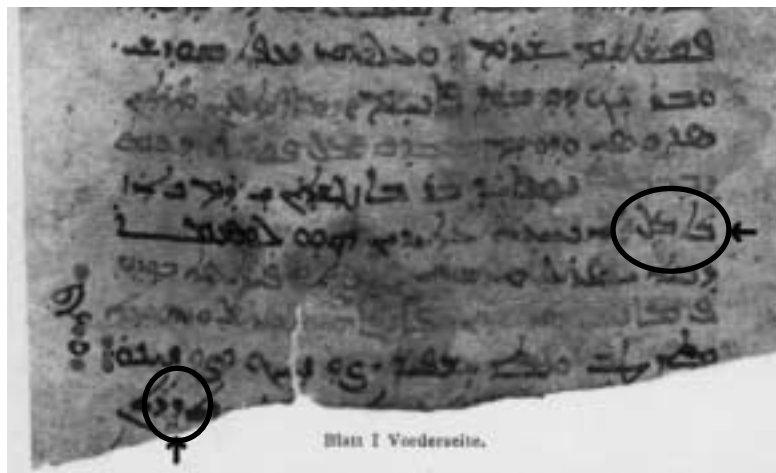


Figure 7. Christian Persian manuscript (Psalms 131-132), showing the letters BHETH and DHALATH: from F. W. K. Müller, "Ein syrisch-neupersisches Psalmenbruchstück aus Chinesisch-Turkistan" in G. Weil (ed.), *Festschrift Eduard Sachau zum siebzigsten Geburtstag gewidmet*, Berlin 1915, Taf. II. The horizontal arrow points to the word for "Babylon", spelled BETH-ALAPH-BHETH-LAMADH (plus some vowel points), demonstrating the difference between BETH and BHETH.