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

L2/03-149

Doc Type: **Working Group Document**
Title: **Final proposal to encode the Old Persian script in the UCS**
Source: **Michael Everson**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2003-05-06**

Old Persian is found in a number of inscriptions in the Old Persian language dating from the Achaemenid Empire. It is an alphabetic writing system with some syllabic aspects. While the shapes of some Old Persian letters may look similar to signs in Sumero-Akkadian Cuneiform, it is clear that only one of them, ⟨LA⟩, was borrowed from Sumero-Akkadian Cuneiform ⟨LA⟩ – and that because *la* is a foreign sign used to represent a sound not used in the Old Persian language. Scholars today agree that the character inventory of Old Persian was newly-invented for the purpose of providing monumental inscriptions of the Achaemenid king, Darius I, by about 525 BCE.

Old Persian is written from left to right. The repertoire contains 36 signs which represent consonants, vowels or sequences of single consonants plus vowels, a set of five numbers, one word divider, and eight ideograms. It is unlikely that more than the characters known today will ever be discovered.

The attested numbers are built up by adding the base numbers together cumulatively:

𐏑 1, 𐏒 2, 𐏓 5, 𐏔 7, 𐏕 8, 𐏖 9

𐏗 10, 𐏘 12, 𐏙 13, 𐏚 14, 𐏛 15, 𐏜 18, 𐏝 19

𐏞 20, 𐏟 22, 𐏠 23, 𐏡 25, 𐏢 26, 𐏣 27

𐏤 40, 𐏥 60, 𐏦 120.

The signs AURAMAZDAA and AURAMAZDAA-2 and the signs DAHYAAUSH and DAHYAAUSH-2 could be seen to be glyph variants, but their conventional attestation in the corpus of Old Persian texts is quite limited and scholars consider it more advantageous to distinguish the forms in plain text representation than to attempt to distinguish them via some other rich-text or VARIATION SELECTOR mechanism.

A. Administrative

1. Title
Final proposal to encode the Old Persian script in the UCS.
2. Requester's name
Michael Everson
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2003-05-06
5. Requester's reference (if applicable)
N1639, N2545.
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)
Yes.
Proposed name of script
Old Persian.
1b. The proposal is for addition of character(s) to an existing block
No.
1b. Name of the existing block
2. Number of characters in proposal
50
3. Proposed category (see section II, Character Categories)
Category B.1.
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?
No combining characters.
4c. If YES, reference
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.
6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson, Fontographer.
7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
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/UnicodeCharacter Database.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
Spacing letter, category "Lu" (uppercase), bidi category "L" (strong left to right)
   1xxA0-1xxC3, 1xxC8-1xxCF
Spacing digit, category "No", bidi category "L" (strong left to right)
   1xxD1-1xxD5
Space separator, category "Zs", bidi category "L" (strong left to right)
   1xxD0

C. Technical -- Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
Yes, N1639.

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?

Günter Schweiger, Petr Vavroušek, and the TITUS project in Frankfurt.

2c. If YES, available relevant documents

http://titus.fkidg1.uni-frankfurt.de/unicode/iranian/3tagung.htm

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

Iranianists, Indo-Europeanists, scholars relating to Ancient Near Eastern Studies, and other scholars.

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

Writing the Old Persian language is not very common these days.

4b. Reference

Schmitt, Rüdiger. 1991. "The Bisitun Inscriptions of Darius the Great. Old Persian Text", in *Corpus Inscriptionum Iranicarum*, London.

Schweiger, Günter. 1998. *Kritische Neuedition der achaemenidischen Keilinschriften (in zwei Bänden) . Band 1: Textband. Transkribierter Text und Übersetzung; Band 2: Katalog. Archäologische Gegebenheiten und kritische Lesungen in Transliteration*. Taimering: Schweiger VWT-Verlag, 1998.

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

Yes.

5b. If YES, where?

Scholarly publications, text editions.

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?

No. Proposed allocation is U+103A0-U+103DF

6b. If YES, is a rationale provided?

6c. If YES, reference

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

Yes.

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?

The name and titles of Darius the Great as written in Persian, Susian, and Babylonian, upon the Rock of Behistun.

I. PERSIAN TEXT.

1.

2.

3.

4.

5.

6.

Sample text from L. W. King, 1901, *Assyrian language: Easy lessons in the cuneiform inscriptions*, London: Kegan Paul, Trench, Trübner & Co. The text reads:

¶ a-da-ma ¶ da-a-ra-y-va-u-š ¶ x-š-a-y-θ-i-y ¶ va-z-ra-ka

¶ x-š-a-y-θ-i-y ¶ x-š-a-y-θ-i-y-a-na-a-ma

¶ vi-i-š-ta-a-s-p-h-y-a ¶ p-u-ça ¶ h-x-a-ma-na-i-š-i-y ¶

Adam, Dārayava[h]uš, xšāyaθiya vazraka,

xšāyaθiya xšāyaθiyanām,

Vištaspahya puça, Haxamanišiya.

I, Darius, the great king,

the king of kings,

son of Hystaspes, the Achaemenian.

TABLE   - Row 1xx: OLD PERSIAN

	1xxA	1xxB	1xxC	1xxD
0	𒀀	𒀁	𒀂	𒀃
1	𒀄	𒀅	𒀆	𒀇
2	𒀈	𒀉	𒀊	𒀋
3	𒀌	𒀍	𒀎	𒀏
4	𒀐	𒀑		𒀒
5	𒀓	𒀔		𒀕
6	𒀖	𒀗		
7	𒀘	𒀙		
8	𒀚	𒀛	𒀜	
9	𒀝	𒀞	𒀟	
A	𒀠	𒀡	𒀢	
B	𒀣	𒀤	𒀥	
C	𒀦	𒀧	𒀨	
D	𒀩	𒀪	𒀫	
E	𒀬	𒀭	𒀮	
F	𒀯	𒀰	𒀱	

G = 00
P = 01

TABLE - Row 1xx: OLD PERSIAN

hex	Name	hex	Name
A0	OLD PERSIAN SIGN A		
A1	OLD PERSIAN SIGN I		
A2	OLD PERSIAN SIGN U		
A3	OLD PERSIAN SIGN KA		
A4	OLD PERSIAN SIGN KU		
A5	OLD PERSIAN SIGN GA		
A6	OLD PERSIAN SIGN GU		
A7	OLD PERSIAN SIGN XA		
A8	OLD PERSIAN SIGN CA		
A9	OLD PERSIAN SIGN JA		
AA	OLD PERSIAN SIGN JI		
AB	OLD PERSIAN SIGN TA		
AC	OLD PERSIAN SIGN TU		
AD	OLD PERSIAN SIGN DA		
AE	OLD PERSIAN SIGN DI		
AF	OLD PERSIAN SIGN DU		
B0	OLD PERSIAN SIGN THA		
B1	OLD PERSIAN SIGN PA		
B2	OLD PERSIAN SIGN BA		
B3	OLD PERSIAN SIGN FA		
B4	OLD PERSIAN SIGN NA		
B5	OLD PERSIAN SIGN NU		
B6	OLD PERSIAN SIGN MA		
B7	OLD PERSIAN SIGN MI		
B8	OLD PERSIAN SIGN MU		
B9	OLD PERSIAN SIGN YA		
BA	OLD PERSIAN SIGN VA		
BB	OLD PERSIAN SIGN VI		
BC	OLD PERSIAN SIGN RA		
BD	OLD PERSIAN SIGN RU		
BE	OLD PERSIAN SIGN LA		
BF	OLD PERSIAN SIGN SA		
C0	OLD PERSIAN SIGN ZA		
C1	OLD PERSIAN SIGN SHA		
C2	OLD PERSIAN SIGN SSA		
C3	OLD PERSIAN SIGN HA		
C4	(This position shall not be used)		
C5	(This position shall not be used)		
C6	(This position shall not be used)		
C7	(This position shall not be used)		
C8	OLD PERSIAN SIGN AURAMAZDAA		
C9	OLD PERSIAN SIGN AURAMAZDAA-2		
CA	OLD PERSIAN SIGN AURAMAZDAAHA		
CB	OLD PERSIAN SIGN XSHAAYATHIYA		
CC	OLD PERSIAN SIGN DAHYAAUSH		
CD	OLD PERSIAN SIGN DAHYAAUSH-2		
CE	OLD PERSIAN SIGN BAGA		
CF	OLD PERSIAN SIGN BUUMISH		
D0	OLD PERSIAN WORD DIVIDER		
D1	OLD PERSIAN NUMBER ONE		
D2	OLD PERSIAN NUMBER TWO		
D3	OLD PERSIAN NUMBER TEN		
D4	OLD PERSIAN NUMBER TWENTY		
D5	OLD PERSIAN NUMBER HUNDRED		
D6	(This position shall not be used)		
D7	(This position shall not be used)		
D8	(This position shall not be used)		
D9	(This position shall not be used)		
DA	(This position shall not be used)		
DB	(This position shall not be used)		
DC	(This position shall not be used)		
DD	(This position shall not be used)		
DE	(This position shall not be used)		
DF	(This position shall not be used)		