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

L2/03-072

**Doc Type: Working Group Document**
**Title:      Revised proposal to encode the Avestan and Pahlavi script in the UCS**
**Source:   Michael Everson**
**Status:   Individual Contribution**
**Action:   For consideration by JTC1/SC2/WG2 and UTC**
**Date:      2002-12-04**

# A. Administrative

**1. Title**
Revised proposal to encode the Avestan and Pahlavi script in the UCS.
**2. Requester's name**
Michael Everson
**3. Requester type (Member body/Liaison/Individual contribution)**
Individual contribution.
**4. Submission date**
2002-12-04
**5. Requester's reference (if applicable)**
N1684
**6. Choose one of the following:**
**6a. This is a complete proposal**
No.
**6b. More information will be provided later**
Yes.

# 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**
Avestan and Pahlavi.
**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**
87
**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 3.
**4b. Is a rationale provided for the choice?**
Uses 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.

**1**

**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?**

No.

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

No.

**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.**

## C. Technical -- Justification

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

Yes. N1684.

**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?**

Jost Gippert, Desmond Durkin-Meisterernst, 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?**

Zoroastrians, Iranianists and other scholars.

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

Used liturgically and by scholars.

**4b. Reference**

P. Oktor Skjærvø in Daniels & Bright 1996.

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

Yes.

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

Scholarly publications.

**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. Proposed allocation is U+10B00-U+10B2F

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

**6c. If YES, reference**

Accordance with the Roadmap. Avestan is used in modern Zoroastrian religion.

**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**

**2**

**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)?**
Yes.
**11b. If YES, is a rationale for such use provided?**
No.
**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 Avestan script was derived as a rationalization and improvement on the original Pahlavi script, which itself was derived from a variety of Aramaic. The Avestans used modified letterforms to distinguish between signs which in Pahlavi had fallen together. Indeed, while Pahlavi words are commonly used within Avestan texts, it was often the case that the scribes had no idea what the values of the words were, because so many of the letters' shapes were identical. The notion of *archigrapheme* has been found to be useful in proposing several of the Pahlavi extension characters: it is often quite clear that the scribes themselves were writing "a bowl and an ear" even though no such letter exists in Pahlavi – the encoding of these enables scholars to represent the same kinds of ambiguities which the scribes themselves were writing. This is not the same as encoding palaeographical variants.

In this encoding, Avestan is taken as the primary form. Letters unique to Pahlavi, or letters with a glyph shape unrelated to the normalized Avestan letters, are encoded separately. In addition to this Pahlavi-specific the combining marks and archigraphemes are added.

Both scripts are written from right to left. Pahlavi is a connected script with some Arabic-like shaping behaviour; Avestan letters are written separately, or touch accidentally in normal cursive hands.

**3**

## TABLE XXX - Row XX: AVESTAN AND PAHLAVI

|   | xx0 | xx1 | xx2 | xx3 | xx4 | xx5 |
|---|-----|-----|-----|-----|-----|-----|
| 0 | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| 1 | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| 2 | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| 3 | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| 4 | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| 5 | ‎ | ‎ | ‎ | | ‎ | ‎ |
| 6 | ‎ | ‎ | ‎ | | ‎ | ‎ |
| 7 | ‎ | ‎ | ‎ | | ‎ | ‎ |
| 8 | ‎ | ‎ | ‎ | | | ‎ |
| 9 | ‎ | ‎ | ‎ | | | ‎ |
| A | ‎ | ‎ | ‎ | | ‎ | ‎ |
| B | ‎ | ‎ | ‎ | | ‎ | ‎ |
| C | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| D | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| E | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |
| F | ‎ | ‎ | ‎ | ‎ | ‎ | ‎ |

**G = 00**
**P = 00**

## TABLE XXX - Row XX: AVESTAN AND PAHLAVI

| hex | Name | | hex | Name |
|-----|------|---|-----|------|
| 00 | AVESTAN LETTER A | | 59 | AVESTAN LETTER PAHLAVI QA |
| 01 | AVESTAN LETTER AA | | 5A | AVESTAN ARCHIGRAPHEME PAHLAVI PA |
| 02 | AVESTAN LETTER AO | | 5B | AVESTAN ARCHIGRAPHEME PAHLAVI MA |
| 03 | AVESTAN LETTER AAO | | 5C | AVESTAN ARCHIGRAPHEME PAHLAVI YA |
| 04 | AVESTAN LETTER AN | | 5D | AVESTAN ARCHIGRAPHEME PAHLAVI EAR |
| 05 | AVESTAN LETTER AEN | | 5E | AVESTAN ARCHIGRAPHEME PAHLAVI ELBOW |
| 06 | AVESTAN LETTER AE | | 5F | AVESTAN ARCHIGRAPHEME PAHLAVI BELLY |
| 07 | AVESTAN LETTER AEE | | | |
| 08 | AVESTAN LETTER E | | | |
| 09 | AVESTAN LETTER EE | | | |
| 0A | AVESTAN LETTER O | | | |
| 0B | AVESTAN LETTER OO | | | |
| 0C | AVESTAN LETTER I | | | |
| 0D | AVESTAN LETTER II | | | |
| 0E | AVESTAN LETTER U | | | |
| 0F | AVESTAN LETTER UU | | | |
| 10 | AVESTAN LETTER KA | | | |
| 11 | AVESTAN LETTER XA | | | |
| 12 | AVESTAN LETTER XYA | | | |
| 13 | AVESTAN LETTER XVA | | | |
| 14 | AVESTAN LETTER GA | | | |
| 15 | AVESTAN LETTER GYA | | | |
| 16 | AVESTAN LETTER GHA | | | |
| 17 | AVESTAN LETTER CA | | | |
| 18 | AVESTAN LETTER JA | | | |
| 19 | AVESTAN LETTER TA | | | |
| 1A | AVESTAN LETTER THA | | | |
| 1B | AVESTAN LETTER DA | | | |
| 1C | AVESTAN LETTER DHA | | | |
| 1D | AVESTAN LETTER TTA | | | |
| 1E | AVESTAN LETTER PA | | | |
| 1F | AVESTAN LETTER FA | | | |
| 20 | AVESTAN LETTER BA | | | |
| 21 | AVESTAN LETTER WA | | | |
| 22 | AVESTAN LETTER NGA | | | |
| 23 | AVESTAN LETTER NGYA | | | |
| 24 | AVESTAN LETTER NGVA | | | |
| 25 | AVESTAN LETTER NA | | | |
| 26 | AVESTAN LETTER NYA | | | |
| 27 | AVESTAN LETTER NNA | | | |
| 28 | AVESTAN LETTER MA | | | |
| 29 | AVESTAN LETTER MYA | | | |
| 2A | AVESTAN LETTER YYA | | | |
| 2B | AVESTAN LETTER YA | | | |
| 2C | AVESTAN LETTER VA | | | |
| 2D | AVESTAN LETTER RA | | | |
| 2E | AVESTAN LETTER SA | | | |
| 2F | AVESTAN LETTER ZA | | | |
| 30 | AVESTAN LETTER SHA | | | |
| 31 | AVESTAN LETTER ZHA | | | |
| 32 | AVESTAN LETTER SHYA | | | |
| 33 | AVESTAN LETTER SHHA | | | |
| 34 | AVESTAN LETTER HA | | | |
| 35 | (This position shall not be used) | | | |
| 36 | (This position shall not be used) | | | |
| 37 | (This position shall not be used) | | | |
| 38 | (This position shall not be used) | | | |
| 39 | (This position shall not be used) | | | |
| 3A | (This position shall not be used) | | | |
| 3B | (This position shall not be used) | | | |
| 3C | AVESTAN PUNCTUATION END OF WORD | | | |
| 3D | AVESTAN PUNCTUATION END OF SENTENCE | | | |
| 3E | AVESTAN PUNCTUATION END OF VERSE | | | |
| 3F | AVESTAN PUNCTUATION FLEURON | | | |
| 40 | AVESTAN PAHLAVI NUMBER ONE | | | |
| 41 | AVESTAN PAHLAVI NUMBER TWO | | | |
| 42 | AVESTAN PAHLAVI NUMBER THREE | | | |
| 43 | AVESTAN PAHLAVI NUMBER FOUR | | | |
| 44 | AVESTAN PAHLAVI NUMBER TEN | | | |
| 45 | AVESTAN PAHLAVI NUMBER TWENTY | | | |
| 46 | AVESTAN PAHLAVI NUMBER FORTY | | | |
| 47 | AVESTAN PAHLAVI NUMBER ONE THOUSAND | | | |
| 48 | (This position shall not be used) | | | |
| 49 | (This position shall not be used) | | | |
| 4A | AVESTAN COMBINING PAHLAVI G | | | |
| 4B | AVESTAN COMBINING PAHLAVI D | | | |
| 4C | AVESTAN COMBINING PAHLAVI Y | | | |
| 4D | AVESTAN COMBINING PAHLAVI S | | | |
| 4E | AVESTAN COMBINING PAHLAVI SH | | | |
| 4F | AVESTAN PAHLAVI KASHIDA | | | |
| 50 | AVESTAN LETTER PAHLAVI HA | | | |
| 51 | AVESTAN LETTER PAHLAVI GA | | | |
| 52 | AVESTAN LETTER PAHLAVI DA | | | |
| 53 | AVESTAN LETTER PAHLAVI HE | | | |
| 54 | AVESTAN LETTER PAHLAVI WA | | | |
| 55 | AVESTAN LETTER PAHLAVI AYIN | | | |
| 56 | AVESTAN LETTER PAHLAVI RA | | | |
| 57 | AVESTAN LETTER PAHLAVI FINAL STROKE | | | |
| 58 | AVESTAN LETTER PAHLAVI LLA | | | |