TO: UTC
FROM: Deborah Anderson, Ken Whistler, Roozbeh Pournader, Andrew Glass, and Laurentiu Iancu
SUBJECT: Recommendations to UTC #147 May 2016 on Script Proposals
DATE: 6 May 2016

The script ad hoc group met three times since the January UTC in order to review a long list of proposals, several of which are very detailed. Nevertheless, only a limited number of documents were able to be discussed before UTC #147. Many preliminary proposals have not yet been reviewed, but can be taken up during the May 2016 UTC.

EUROPE
1. *Latin*
Document: L2/16-125 Proposal to add Medievalist punctuation characters – Everson

Discussion: We reviewed this document, which is a revised version of the document seen at the last UTC, but with changes in the text based on comments over the Unicore email list. It also includes a new figure with an example of MEDIEVAL COMMA, as had been requested in the earlier script recommendations. (The two characters approved at the January 2016 meeting have been removed.)

The following points were raised in discussion:
- PARAGRAPHUS appears to be a good candidate for encoding. Add figure 27, which contains an elaborate glyph for PARAGRAPHUS, to the list of figures for this character. Discuss the range of glyphs (four are shown, in figs. 1, 2, 21, and 27), and provide justification for the choice of the representative glyph. Should two marks be encoded?
- PUNCTUS ELEVATUS MARK is also a good candidate. Provide samples of the glyphs and discuss their range. Which glyph should be chosen for modern text? For example, the proposed glyph (below, left) contrasts with check-mark in figure 10 (middle, with the two elements directly on top of one another) and fig. 16 (right). The middle and right samples have elongated tails.
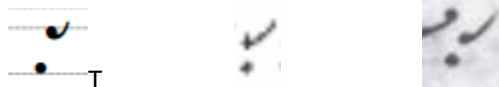
- Strong evidence for MEDIEVAL COMMA is now provided by figure 27, and hence this character in our view is a third candidate for encoding.

Comments on other characters or figures:
- The caption for Figure 7 needs to be adjusted. The first few characters identified with arrows don't seem to match the names in the caption.

- The evidence for the POSITURA MARK all comes from Parkes. Provide examples from manuscripts not in Parkes.

- TWO DOTS OVER COMMA is only shown in transcription (figures 5 and 9) with no manuscript examples. This character appears to be adequately handled as a sequence of <MIDDLE DOT, COMMA, MIDDLE DOT>.

- The examples of COLON WITH SIDEWAYS REVERSED RAISED COMMA demonstrate that the data has been mis-analyzed, and should not be handled as a single unit. What is the second element, and how should it be represented?

  The proposal suggests that figure 6 is the basic form and figure 8 is a variant, which causes a problem when representing text as it appears in the manuscripts.

  Cf.

  figure 6: vs. transcription: and the variant shown in fig. 8: and

   vs. transcription
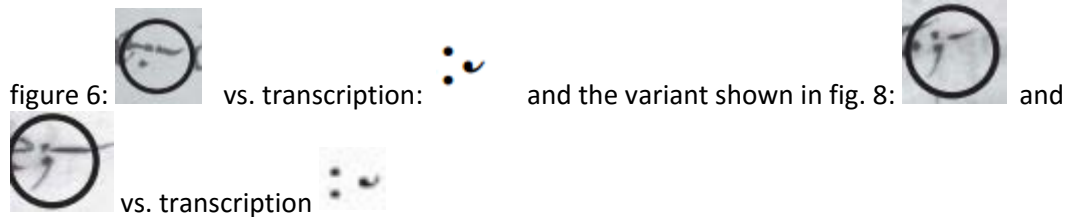
- PUNCTUS INTERROGATIVUS MARK is transcribed as "?" in figure 13, but the other examples (except figure 7) all come from Parkes. Provide evidence from other sources.

- PUNCTUS EXCLAMATIVUS
    - Get verification that Parkes' example in (c) below is PUNCTUS EXCLAMATIVUS.
    - The examples of PUNCTUS EXCLAMATIVUS in figures 17 and 18 (below, [a] and [b]) don't clearly show two dots below (with space between the upper dot and the bar), compared to Parkes' transcription (c) and the proposed glyph (d):

  (a)   (b)   (c)   (d)
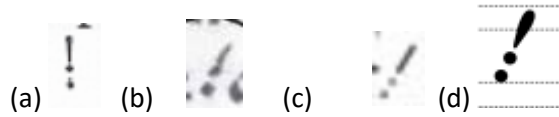
  Unless another manuscript with clear examples showing two separate dots can be provided, the modern ! U+0021 should be used to represent this character, which itself ultimately derives from PUNCTUS EXCLAMATIVUS.

- No additional evidence for VERTICAL FIVE DOTS from a manuscript or inscription has been provided, other than figure 26 (where the Latin transcription does not capture the four-dot or the five-dot character) and a TEX query. Provide more evidence that VERTICAL FIVE DOTS is used in contrast, i.e., showing systematic representation of the five dots in Runic transcriptions where the four vertical dot and five vertical dot marks are distinguished. Include this character in a proposal once additional examples can be provided.

Recommendations: We recommend the UTC invite the author to:
1. submit a final proposal for the following three characters (with the appropriate illustrations): MEDIEVAL COMMA, PARAGRAPHUS, and PUNCTUS ELEVATUS, and remove the other, unrelated characters.
2. submit a preliminary proposal that addresses the remaining suggested candidates. Discuss why Parkes is given such emphasis, and invite input from other experts on the proposal.

2. Georgian
Document: L2/16-081 Revised proposal for the addition of Georgian characters – Everson

Discussion: We reviewed this proposal, which requests Georgian be changed to a bicameral script, and proposes a new block of capital letters to the standard (Mtavruli). This revised proposal includes text discussing alternatives (§11, pp. 5-11), as well as several letters of support from font foundries and governmental offices.

A number of points were raised in the discussion:
- What are the schools in Georgia currently teaching?
- Can the author get engineering implementers to provide their point of view on the change, and discuss what impact it would have?

Recommendation: We recommend the UTC discuss this proposal, and encourage the author and UTC members engage with IT members in Georgia so they are involved in the discussion.

MIDDLE EAST
3. Hebrew
Document: L2/16-086 Proposal of an additional character to the Hebrew Unicodes - Sheva Na – Eliezer May

Discussion: We discussed this document, which requests a bold HEBREW POINT SHEVA character be encoded in order to indicate whether SHEVA is vocalized or not, since the pronunciation is not predictable. The author points to the implementation by the company Davka, whose fonts contains the glyph.

The requested character appears to be the way one company approaches the problem. If the character had the general backing of scholars, industry, and/or the Israeli National Body, there would be a stronger case for standardizing it.

Recommendations: We recommend the UTC discuss the proposal, and forward feedback to the proposal author.

AFRICA
4. Egyptian hieroglyphs
a. Control characters
Documents: L2/16-090 Comments on three control characters for Egyptian Hieroglyphs – Nederhof, Rajan, et al.
L2/16-104 Observations: L2/16-090 [Egyptian] – Richmond
Reference doc: L2/16-018 Proposal to encode three control characters for Egyptian Hieroglyphs (revised) – Glass and Richmond

Discussion: We reviewed L2/16-090, which contains three responses to the three Egyptian Hieroglyph control characters that were discussed and approved at the January 2016 UTC. A response to L2/16-090 was written by Bob Richmond in L2/16-104..

In L2/16-090, the respondents did not object to the HORIZONTAL JOINER and VERTICAL JOINER, but found LIGATURE JOINER to be a problematical character.

Discussion on L2/16-090 focused on three issues, and comments are noted below:

1. Reading order of the characters
For complex writing systems such as Egyptian hieroglyphs or Japanese, the visual order of characters should be separate from phonetic order. In complex writing systems, combining visual and phonetic order can result in visual ambiguity, where one reading could have two different visual sequences. One could, however, keep a separate field (or use mark-up) with the phonetic reading, which could be carried along with the data.

2. Superposition of characters
As noted by Bob Richmond on the bottom on page 1 of L2/16-104, superimposed characters, or "monograms", were handled as atomic units in the original Unicode proposal (L2/07-197), MdC, and Hieroglyphica. How many characters in the new proposal L2/16-079 contain such characters, and should they be handled as pieces with a new joiner? The users and Unicode proposer should discuss this, and come to consensus on what to do.

3. Unpredictable ligatures
Because of the questions Nederhof raises on the bottom of page 3 and top of page 4 regarding how G29&N29 (etc.) are rendered, we advise the authors of L2/16-018 prepare a list of ligatures identifying what "A&B" is supposed to represent, in order to guarantee predictable rendering of the combinations.

Other comments:
- Nederhof, page 1, footnote 1, mentions that Michel Suignard's proposal L2/16-028 is lacking any suggestion on how to determine where the proposed signs were found, their meaning, etc. The authors should now refer to the latest proposal from Michel Suignard, L2/16-079 ("Preliminary draft for the encoding of an extended Egyptian Hieroglyphs repertoire"), especially page 4, where a database design is described that will provide source data for each character, as is done for the Han characters in the Unihan database.

- Nederhof, page 7, cites the Ideographic Description Characters as "precedents for control characters that define spatial arrangements of characters". These characters are not control characters used to construct glyphs on the fly, but are only to provide rough descriptions (see section 18.2 of the Unicode Standard 8.0). Full glyph description would require, instead, a glyph description language, such as that required for Han ideographs. See, for example, CDL: www.wenlin.com/cdl .

- The proposal for the three characters (L2/16-018) showed examples on pages 3ff. of the control characters working with off-the-shelf software and fonts (i.e., with standard OpenType font features and an operating system with the Universal Shaping Engine). However, users are under no obligation to use the control characters, and developers can create their own software.

Recommendations: We recommend the UTC review this document and discuss it. We suggest the authors of L2/16-018 prepare a list of ligatures and discuss how to handle "monograms".

b. Extended Egyptian Hieroglyphs
Document: L2/16-079 Preliminary draft for the encoding of an extended Egyptian Hieroglyphs repertoire – Suignard

Discussion: We quickly reviewed this preliminary proposal, which represents the full repertoire of additions to Egyptian Hieroglyphs, based primarily on Hieroglyphica.

This document builds off earlier proposals, which only contained a portion of the entire set (L2/16-028 covered characters in categories A [Man and his occupations] through F [Parts of mammals] and L2/15-240 only contained characters from the A set). This proposal now includes a revised introduction, and describes the creation of a database with source information.

Recommendations: We recommend the UTC review this proposal at their leisure.

5. Arabic
Al-Dani Quranic Marks (in Libya)
Documents: L2/16-102 Consolidated Comments by Mansour, Evans, and Abudena on L2/16-056 Al-Dani Quranic Marks – Anderson
L2/16-100 Comments on L2/16-056 Proposal to encode AlDani Quranic Marks – Mussa A. A. Abudena
Reference docs:
L2/16-056 Proposal to encode Al-Dani Quranic marks used in Quran published in Libya – Lateef Sagar Shaikh
L2/16-044 Proposal to encode Quranic marks used in Quran published in Libya with Commentary – Mussa A. A. Abudena

Discussion: We reviewed the documents on the Al-Dani Quranic marks. The following summarizes the document trail:
- Mussa Abudena wrote the initial proposal (L2/15-329) and revised it in 2016 (L2/16-044), identifying the specific characters in the figures.
- Lateef Shaikh submitted a proposal (L2/16-056) drawing heavily from the work done by Mussa Abudena (L2/16-044)
- Kamal Mansour and Lorna Evans sent comments on L2/16-056 over email, and Mussa Abudena submitted a separate document (L2/16-100); all the comments from Mansour, Evans, and Abudena to L2/16-056 were then collected and included into a single document, L2/16-102.

The script ad hoc concentrated on the "Consolidated Comments by Mansour, Evans, and Abudena" document (L2/16-102).

The full proposal by Mussa Abudena and Lateef Shaikh have not been thoroughly analyzed, and there may be other good candidates for encoding, such as #30 and #31 in L2/16-044 and the black-dotted character in L2/16-056 on the bottom of page 9 and top of page 10.

Specific comments based on discussion of part I of L2/16-102 (listed by character numbers in L2/16-044):

#1 ◌ ARABIC ALTERNATE DAMMA and #2 ◌ ARABIC ALTERNATE DAMMATAN are variants of damma (#1) and dammatan (#2) or open dammatan.

#3 ◌ ARABIC ALTERNATE DAMMATAN 2 is a sequence of damma and U+06E2 meem.

#4 ◌ ARABIC ALTERNATE FATHATAN is a fatha and U+06E2 meem.

#5 ◌ ARABIC ALTERNATE KASRATAN is a kasra and low meem U+06ED.

#6 ◌ ARABIC ALEF MOKHASES is not a dagger alef, but does appear to be eligible for encoding, although more research is required.

#7 ◌ARABIC LETTER ALEF MOKHASES WITH MADDAH ABOVE is #6 with maddah above (U+0653) or   small high madda (U+06E4)

#8 ◌ ARABIC ALTERNATE SIGN OF WAQAF is a legitimate character to encode, but a name is needed, as it is not a wasla.

#9 ◌ ARABIC SMALL HIGH YEH WITH FATHA ABOVE also appears to be an eligible character, but more investigation is needed to understand how it is used, its meaning, and pronunciation.

#10 ◌ ARABIC SMALL WAW BELOW is a valid character, and should be encoded.

#11 ◌ ARABIC LETTER HIGH YEH MAQUSA WITH DOT ABOVE is an alternate version of U+06E8 arabic small high noon, used only once in the Koran. There is no need to encode this character.

#12 ◌ ARABIC LETTER SUKUN BELOW could be an Arabic small low rounded zero, to be paired with U+06DF ◌ small high rounded zero. (The usual sukun in the Quran is U+06E1 ◌ small high dotless head of khah.)   When it appears under alef (and perhaps under other characters), it may be a separate character, related to #15. In sum, #12 requires further study.

#13 ﹔ ARABIC SMALL WAW INSIDE TATWEEL and #14 ﹔ ARABIC HAMZA INSIDE TATWEEL are ligatures of tatweel and small waw above (in character #13) and tatweel with hamza (in character #14). In other Qurans, the waw and hamza can be placed above the tatweel. In sum, these graphic variants of sequences should not be separately encoded.

#15 ﺍ ARABIC ALTERNATE ALEF WASLA, like other characters that appear combined with solid dots and circles, will require further research. Compare the following, for example:

What are the dots?

#16 ✒ ARABIC ALEF MOTHAFAR is a ligature of a lam and superscript alef, which would be difficult to implement in a font. Encoding it would not help in the representation of text.

Comments in response to part II of L2/16-102 (from Mussa Abudena):
a) YA
[Abudena:] "This shape of (Ya) (See Yeh with small higher yeh page 9) not included in Al-dani at all"

This comment needs further clarification from the author.

b) SAD and DAD WITHOUT SENNA
[Abudena:] "[T]he statement on page 6 [of L2/16-056], 'Which means that removing senna is a calligraphic style and should be programmed in the font if required' is incorrect."

Explain the difference in meaning between the two forms below. Did the scribe just miss a stroke? Further information is needed to justify separate characters. What are the rules for their use?

 (Abudena #10, fig. 6, L2/16-044)   (Abudena #11, fig. 7, L2/16-044)

c) ALEF KHINGARIH
Regarding the following statement from by Shaikh (L2/16-056):

> [Alef Khingarih ۟ on page 6 of L/216-056] has same visual and linguistic properties as Arabic Letter Superscript Alef – U 670. Therefore this should not be encoded as a separate character."

Abudena responded that Alef Khingarih has a difference in size in Aldani, as shown in the following chart:

Comparison between the writing of Quran in narration Hafs (Khrraz) and narration Qaloon (Aldani)

No.	Name	Khrraz	Aldani
7	ARABIC ALEF KHINGARIH	۱	۱

In our view, the larger size reflects the style of writing in the different Qurans, and does not appear to be a separate character.

d) ALEF WASLA

[Abudena:] "The original ALDANI ARABIC ALEF WASLA is ALEF without dot over it (۱), doesn't as in noted on page 2 of L2/16-056 (mark No. 15), then dot, Sukun, kasra, or fatha two of them added to it.

The comment needs further clarification from the author.

e) ALEF WASLA
See discussion under #15 above.

f) HAMZA below LAM_ALEF
[Abudena:] "Incorrect position of HAMZA below LAM_ALEF in page 9 [of L2/16-056]"

It appears that the figures in L2/16-056 may be incorrect; the character sequence needs to be checked.

g) DAMMA
Abudena comments on the shape of the damma above an alef.

This comment is part of the larger conversation mentioned under #15.

h) DAMMATAN IN CASE OF ETHHAR
Abudena reports his character #2, DAMMATAN IN CASE OF ETHHAR, was omitted in L2/16-056.

This character is probably the ◌ͮ open dammatan (U+0841).

I) YEH WITH HIGH YEH MAQUA ABOVE

Abudena notes his proposed character #15 ؈ YEH WITH HIGH YEH MAQUA ABOVE was also left out by Shaikh in L2/16-056.

This character appears to be a sequence of <U+064A yeh, U+06E7 high yeh, U+0651 shadda>.

Recommendations: We recommend the UTC approve #8 ARABIC ALTERNATE SIGN OF WAQAF, after deciding upon a more accurate name, and #10 ARABIC SMALL WAW BELOW. Several of the other proposed characters will require further study (i.e., #6, #9, #12, and #15), and we recommend UTC members with expertise in Arabic be encouraged to participate in the review of such characters.

6. Medefaidrin
Document: L2/16-101 Medefaidrin (revised) – Rovenchak et al.

Discussion: We reviewed this proposal which has gone through a series of revisions (L2/15-117, L2/15-298, L2/16-020), based on feedback from the UTC. The latest proposal, L2/16-101, answers the outstanding questions that remained.

If it is possible, provide another example of the EXCLAMATION MARK, besides the handwritten example on page 13 (which was added in the latest version of the proposal).

Recommendation: We recommend the UTC review this proposal and approve encoding this script.

7. Mandombe
Document: L2/16-077 Proposal for encoding the Mandombe script in the SMP of the UCS – Rovenchak et al.

Discussion: We reviewed this revised proposal, which now includes the names of the characters written out and additional details on the characters in the script.

The following are comments on the proposal, raised during discussion:
- Provide a list showing which syllables take combining marks, and their placement.
- Various arrows are proposed for minus, plus, equals, alternate colon, and semicolon. Could existing arrows be used?
- Describe how the four acute accents are used.
- Are there any instances of syllables with a diphthong and post-nasalization? How would LUAN be written?
- Why are there gaps in the chart? Are other characters expected?
- Adjust the spelling from "diphtong" to "diphthong"

Recommendation: We recommend the UTC members review this proposal and send feedback (including the comments above) to the authors.

8. Garay
Document: L2/16-069 Proposal for encoding the Garay script in the SMP – Everson

Discussion: We reviewed this proposal which is a revision of an older proposal from 2012 (L2/12-139, which builds off L2/11-181). The latest proposal addresses questions that had arisen during discussion of the 2012 proposal (such as the "tail" of final letters), and adds more information on the vowel signs and the figures. It includes transcriptions in figures 8 and 9.

In order to confirm the encoding model, we recommend the author provide a set of six different, illustrative Wolof words written in Garay, IPA, and Latin in a revised version of the proposal.

Recommendation: We recommend the UTC members review the proposal and send feedback to the author.

EAST AND CENTRAL ASIA
9. Khitan
Document: L2/16-113 Towards an Encoding of the Khitan Small Script – West et al.

Discussion: We reviewed this lengthy proposal, which contains good documentation on the characters and sources. The proposal also posed specific encoding model questions (§4, pp. 14-17). The relevant background is summarized below.

Khitan Small Script (KSS) is an historic script that is written vertically, with columns running right to left. The script contains logograms, typically written in isolation as single characters, and phonograms, which generally occur in blocks of between 2-8 characters.

Within a block, the phonograms are ordered left-to-right, top-to-bottom, as on left below (with a sample showing layout for 8 characters, from Table 1, p. 9), or in the alternative structures (center and right, from Tables 2 and 3, p. 10):

The proposal offers different approaches for the encoding model, summarized below:

a) encode each attested phonogram block as an atomic precomposed character;

b) encode logographic characters, which cannot combine into blocks, separately from phonographic characters, which automatically combine with adjacent phonograms to form blocks;

c) allow contiguous KSS characters to form a block automatically, but use SPACE (or punctuation mark or other non-KSS character) to demarcate blocks and ZWSP to prevent block formation; encode a format character to identify that the first character of a block should be rendered in the alternative structure (see Tables 2 and 3, p. 10);

d) do not allow blocks to be formed automatically, but use ZWJ between adjacent pairs to indicate they should be rendered as a block;

e) use a script-specific format character to indicate the KSS characters should be rendered as a block in the standard structure (Table 1, p. 9) and a second format character to specify the block should be rendered in the alternative structure (as in Tables 2 and 3, p. 10), placing the block-forming format character before the first character or, alternatively, after it.

Comments raised during discussion:
- We agree that Khitan Small Script should not be unified with CJK.

- The approach of encoding the individual pieces -- Hangul-style -- is correct in our view, but we do not recommend use of SPACE, ZWSP, or ZWJ, since the rendering would be less reliable. Instead, we recommend encoding two Khitan-specific control characters: a horizontal and a vertical stacker. For a comparable case, compare the control characters proposed for the quadrat formation of Egyptian hieroglyph signs (L2/16-018), which has a more complex quadrat model.

  Two Khitan control characters would allow formation of quadrats as in figure 13 (below, left), as well as the linear layout in figure 14 (below, right), which would not use the control characters:

- Can the author verify that there are no examples of multi-row spanning (as below)?

Recommendations: We recommend the UTC review this document and discuss the encoding model. We also suggest the authors revise the encoding model section of the proposal, so it is considerably simplified and presents only the feasible alternatives, and then invite further review of the revised version by other Khitan experts.

SOUTH and SOUTHEAST ASIA
10. Kaithi
Document: L2/16-097 Proposal to encode Kaithi combining enumeration sign – Yang

Discussion: We reviewed this document which proposes one new Kaithi character. The proposal builds off discussion in the original Kaithi proposal by Anshuman Pandey (L2/08-194) and later documents (L2/08-370 and L2/08-400).

The current U+110BD KAITHI NUMBER SIGN is a control format character that can theoretically occur above or below a number, and its position is not predictable. According to the main Kaithi proposal, if the number sign is above the number, it indicates a number is in an itemized list. If it occurs below the sign, it is a numerical reference.

The current proposal suggests that U+110BD KAITHI NUMBER SIGN be used in numerical reference contexts (where the mark appears below the digit), and a new character KAITHI NUMBER SIGN ABOVE be assigned to itemized list usage (where the mark appears above the digit). (The spacing version of the character is already encoded at U+110BC KAITHI ENUMERATION SIGN.)

The proposal seems sound, and we recommend the encoding of this character.

Recommendations: We recommend the UTC approve the character KAITHI NUMBER SIGN ABOVE, but verify the location U+110C2 is acceptable.

11. Cham
Document: L2/16-082 Proposal to encode seven Cham characters in the BMP (WG2 N4711) – Everson

Discussion: We discussed this proposal, which requests the addition of seven characters for Western Cham. Western Cham is used in Cambodia and Vietnam, and Eastern Cham is used in southern Vietnam. The current style for the Cham code chart is Eastern Cham. The Cham proposal (L2/06-257) was based on Eastern Cham materials.

The seven characters appear valid. The following were comments on the proposal:
- Fix the CFL in the Western Cham chart (cf. U+AA2F CHAM VOWEL SIGN O, which appears on the left side in Eastern Cham chart, but on the right side in the Western Cham chart)
- Verify the transliteration for NUN
- Identify the sources of the images
- Are the proposed characters only used in Western Cham? (If so, are Eastern Cham versions needed?)
- Provide a chart showing the letter shapes of Eastern vs. Western Cham
- Provide a chart comparing the final vs non-final forms

Recommendations: We recommend the UTC discuss this document, and encourage interested parties collaborate on a document that outlines reasons for and against the disunification of Western Cham. Once a decision on the disunification is made, the UTC can again take up the potential approval of the seven characters (in a revised proposal, with additional information requested above). We also recommend the current code chart specify that the glyphs are those used in Eastern Cham.

INDONESIA
12. Buginese
Document: L2/16-075 Proposal to encode VIRAMA signs for Buginese – Pandey

Discussion: We reviewed this proposal, which requests two virama characters be added to the Buginese block, filling the last two open holes in the block. These viramas would function as vowel killers. The first proposed character, BUGINESE SIGN VIRAMA-1, is used in the Sumbawa and modern Bugis varieties of the script, and the second character BUGINESE SIGN VIRAMA-2 is used in the Ende and Bima script-varieties.

The proposal gives a sketch of the history of virama as part of Buginese Unicode proposals, though the history extends a bit earlier: the virama was initially proposed by Michael Everson in 1997, based on its appearance in one font where the glyph had the shape of a combining underline (WG2 N1657.htm = L2/98-01). A virama with this shape was carried over into later proposals (L2/99-011 [=N1930] and L2/03-191 [=N2588]). In 2003, Daniel Kai submitted a document (L2/03-253) stating the virama appeared in at least one modern grammar, but had a different shape. He provides the shape ⚛ in L2/03-254, mentions it appeared in a 1991 grammar with this shape, but says a virama does not occur in historical works. In L2/03-320r, Everson proposed the virama with a different shape: a raised dot to the right, based upon more recent fonts. The proposed VIRAMA-1 shape matches that mentioned by Daniel Kai. VIRAMA-2 was not be discussed in earlier proposals.

The evidence seems to be solid, with examples from modern signage, a recent primer, and a web page, but a number of questions were raised:
- Where did the viramas come from, and are they related to one another? Do they co-occur? Are both needed? If so, provide for justification for two viramas.
- What is the relation between these two and the virama discussed in Michael Everson's proposal (L2/03-320R)?
- Several shapes appear: figure 2 and 6 vary from the proposed shapes. How were the representative shapes arrived at?

Recommendation: We recommend the UTC review the proposal, and relay feedback to the proposal author.

SYMBOLS
13. Astrology
a. Astrological Plutos
Document: L2/16-067 Astrological Plutos – Faulks

Discussion: We reviewed this proposal, which requests four astrological symbols for Pluto. The characters are well justified and the proposal is well researched. The author has, in our opinion, correctly distinguished between abstract characters and variants.

The names for planets don't distinguish between "astronomical" or "astrological", with the exception of U+26E2 ASTRONOMICAL SYMBOLS FOR URANUS, which is differentiated from the astrological symbol U+2645 URANUS. PLUTO VARIANT-1, -2, etc., might be one option for the names. The code points (U+2BD4, U+2BD5, U+2BD6, and U+2BD7) are acceptable.

Recommendations: We recommend the UTC approve the four characters for encoding in the Miscellaneous Symbols and Arrows block at the proposed code points, but discuss the most appropriate names.

b. Uranian Astrology Symbols
Document: L2/16-064 Extra Symbols from Uranian Astrology Symbols – Faulks

Discussion: We reviewed this proposal which requests a set of 8 symbols used by astrologers. In our opinion, short names (i.e., "Cupido", "Hades", etc.) should suffice. The proposal otherwise is well-researched, and the code points are acceptable.

Recommendations: We recommend the UTC approve the 8 proposed characters in the Miscellaneous Symbols and Arrows block, after discussing which names would be appropriate. Our suggestion would be:
U+2BE0 CUPIDO
U+2BE1 HADES
U+2BE2 ZEUS
U+2BE3 KRONOS
U+2BE4 APOLLON
U+2BE5 ADMETOS
U+2BE6 VULCANUS
U+2BE7 POSEIDON

c. Additional Symbols for Astrology
**Document:** L2/16-080 Additional Symbols for Astrology – Faulks

Discussion: We reviewed this proposal, which requested 13 characters, drawn from a variety of sources. The proposal contains extensive documentation for each proposed character.

The examples provided for many of the proposed characters came from web pages showing charts generated by various software applications, rather than from printed publications.

The naming patterns proposed, "ASTROLOGICAL PLANET X", "ASTROLOGICAL SYMBOL FOR X", and "ASTROLOGICAL POINT X", don't match the general conventions of character names used elsewhere (cf. U+26B4 PALLAS and U+26B6 VESTA, etc.). We suggest short names, and use subheadings to indicate the characters as astrological signs (cf. subheads before U+2600, U+26D9, and U+26B3).

Specific comments on characters that were raised during discussion:
- TRANSPLUTO is well-documented and a good candidate for encoding.
- PROSERPINA, a Russian "hypothetical" planet that is said to orbit the sun beyond Pluto, appears to be also well-supported by documentation. What relation, if any, is there between the proposed symbol and the one appearing on the 26 Proserpina Wikipedia page, ⚵ (identified as an asteroid)?
- PRIAP and RAZI are less well-known and seem to be not sufficiently documented. Several examples come from software-generated images. Provide more established sources in support of these characters. Also, is the spelling RAZI (p. 10) or RASI (p. 11)?

- HYGIEA and ASTRAEA share the same samples in the proposal, and appear more often published in charts than other main-belt asteroids (outside of Ceres, Pallas, Juno and Vesta [p. 13]). We noted that HYGIEA has a different spelling than U+1F54F BOWL OF HYGIEIA, but HYGIEA appears to be the usual spelling for the asteroid. We noted that HYGIEA's proposed glyph ⚕ varies slightly from 10 Hygiea on Wikipedia (which has the symbol ⚕), and appears to be a variant. In our view, HYGIEA and ASTRAEA are good candidates for encoding.
- PHOLUS and NESSUS were discovered in the 1990s and their astrological symbols devised afterwards in the 1990s. However, these symbols don't yet seem to have well-established usage, and the examples come from screenshots, software, or software manuals.
- WHITE MOON SELENA appears to be well-established. Evidence is provided of it in printed text (7.02). We consider this character eligible for encoding.
- WHITE MOON SELENA WITH RAYS looks to be a variant of WHITE MOON SELENA. Provide additional evidence for the distinction of the two forms.
- BLACK DIAMOND ON CROSS seems to be a good candidate for encoding.
- BLACK CRESCENT WITH BAR and ASTROLOGICAL POINT PRIAPUS require more evidence from established sources.

Recommendations: We recommend the UTC review this proposal and accept the following characters in the Miscellaneous Symbols and Arrows block at the next available code points:
TRANSPLUTO
HYGIEA
ASTRAEA
WHITE MOON SELENA
BLACK DIAMOND ON CROSS
After discussion, the UTC may also decide to accept PROSERPINA.

14. Copyleft Symbol
Document: L2/16-059 Proposal to add the Copyleft Symbol to Unicode – Faulks

Discussion: We reviewed this proposal for the Copyleft symbol. The proposal is well-researched and provides examples in running text. We suggest placing the character in the same column as U+1F12B CIRCLED ITALIC LATIN CAPITAL LETTER C.

Recommendations: We recommend the UTC accept the COPYLEFT SYMBOL, but use the code point U+1F12F in the Enclosed Alphanumeric Supplement block.

NUMBER SYSTEMS
15. Tally Marks
Document: L2/16-065 Two Western-style tally marks - Lunde and Miura

Discussion: We reviewed this proposal, in which the author has expanded the section on Similar Characters (page 2), with clear rationale for encoding the marks, and noted 16 similar-looking TALLY MARK ONE characters (page 3). This proposal has ably answered the questions and comments made during the UTC discussion of the earlier proposal, L2/15-328.

Recommendations: We recommend the UTC approve the two characters U+1D377 TALLY MARK ONE and U+1D378 TALLY MARK FIVE in the Counting Rod Numerals block.

16. Siyaq
a. Number Mark
Document: L2/16-117 Comments on Proposal to Encode Siyaq Number Mark (L2/15-074R) – Afshar

Discussion: We reviewed the Proposal to Encode the SIYAQ NUMBER MARK (L2/15-074R) with Shervin Afshar, who has investigated the mark further, and summarized his findings in L2/16-117. In his view, there is no case for a NUMBER MARK. Its appearance as a uniquely defined Siyaq number mark is from Pihan, where it appears it was created for presentation purposes only. Details of the use of elongation in words or letters alongside Siyaq numbers is provided in L2/16-117.

Recommendation: We recommend the UTC not encode the Siyaq Number Mark based on the research by Shervin Afshar.

b. Diwani/Ottoman Unification
Document: L2/16-116 Comments on Proposal for Diwani/Ottoman Siyaq Unification (L2/15–340) – Afshar
Reference docs: L2/16-017 Proposal to Encode Arabic Siyaq Numbers in Unicode – Pandey
L2/15-340 Unification of 'Diwani' and 'Ottoman' Siyaq Numbers – Pandey

Discussion: We discussed Diwani and Ottoman unification, based on the research by Shervin Afshar (now summarized in L2/16-116). At this point, the case for unifying Ottoman/Diwani is not clear-cut. Ottoman Siyaq is better understood and hence makes a stronger case for encoding. Diwani could be added as a distinct system encoded in a separate block later or, if more evidence is provided, it could be unified with Ottoman (and a mapping between the two provided). Afshar will continue his review and provide detailed comments on Ottoman for a future UTC meeting.

Recommendation: We recommend the UTC members review L2/16-116.