

**TO:** UTC  
**FROM:** Deborah Anderson, Jan Kučera, Ken Whistler, Roozbeh Pournader, and Peter Constable<sup>1</sup>  
**SUBJECT:** Recommendations to UTC #175 April 2023 on Script Proposals  
**DATE:** April 21, 2023

The Script Ad Hoc group met on February 15, March 10, and April 7 2023, order to review proposals, with a special meeting on March 17, 2023 on Seal. The following represents feedback on proposals that were available when the group met.

## Table of Contents

A. PROPOSALS REQUIRING UTC ACTION .....	2
I. MIDDLE EASTERN SCRIPTS .....	2
1 Arabic .....	2
2 Sumero-Akkadian Cuneiform.....	3
II. SCRIPTS FROM EAST ASIA .....	3
3 Khitan Small Script .....	3
4 Tangut .....	4
III. SYMBOLS.....	4
5 Alchemical Symbols .....	4
6 Symbol: FORINT SIGN .....	5
IV. PUBLIC REVIEW FEEDBACK.....	6
7 Arabic .....	6
8 Bima .....	7
9 Sharada .....	8
B. RECOMMENDATIONS FOR 16.0.....	9
C. DOCUMENTS NOT REQUIRING UTC ACTION (by script, in alphabetical order) .....	10
10 Bima .....	10
11 Egyptian Hieroglyphs .....	11
12 Jurchen.....	12
13 Khmer.....	12

---

<sup>1</sup> Also participating were Craig Cummings, Lorna Evans, Manish Goregaokar, Liang Hai, Ned Holbrook, John Hudson, Kushim Jiang, Frank van de Kasteelen, James Kirby, Robin Leroy, Norbert Lindenberg, Steven Loomis, Kamal Mansour, Cheon Hyeong Sim, Michel Suignard, Harald Tveiten, Lawrence Wolf-Sonkin, and Ben Yang. The text for the comments and recommendations was based on notes taken by Debbie Anderson and Jan Kučera.

14 Latin .....	13
14a Initial Teaching Alphabet .....	13
14b Slovene Metelko Alphabet .....	13
15 Seal .....	14
16 Sidetic .....	14
17 Tulu .....	14
18 Other topics: .....	15
18a Do Not Emit .....	15
18b Encoding Proposal Templates .....	16
18c ScriptExtensions .....	16
D. FYI DOCUMENTS .....	17
Arabic .....	17
Badugu .....	17
Old Hungarian .....	17
E. IN PROCESS .....	18
Other script and character topics in process .....	18

## A. PROPOSALS REQUIRING UTC ACTION

### I. MIDDLE EASTERN SCRIPTS

#### 1 Arabic

##### **Action: For UTC discussion and decision**

**Document:** [L2/23-103](#) Proposal for ARABIC BIBLICAL END OF VERSE – Evans

**Recommendation:** We recommend the UTC make the following disposition:

**SAH-UTC175-R1:** Consensus: Provisionally assign U+10ED0 ARABIC BIBLICAL END OF VERSE, with name and glyph as given in L2/23-103. (Reference: Section 1 of L2/23-083)

**Action Item** for Lorna Evans, SAH: Provide a font to Michel Suignard and the Charts group for U+10ED0 ARABIC BIBLICAL END OF VERSE. (Reference: Section 1 of L2/23-083)

**Action Item** for Ken Whistler, UTC: Update the Pipeline to add U+10ED0 ARABIC BIBLICAL END OF VERSE as a provisionally assigned character. (Reference: Section 1 of L2/23-083)

**Note:** The SAH reviewed an earlier version of the proposal; the posted version has accommodated most of the comments below.

**Comments:** We reviewed a proposal for an Arabic character that marks the end of a verse in biblical texts in Urdu. Unification with U+235C APL FUNCTIONAL SYMBOL CIRCLE UNDERBAR was discussed but rejected, based on character identity, as well as the bidirectional properties of the APL block. We consider the character to be ready for encoding, with the following comments:

- The name is too generic, so ARABIC BIBLICAL END OF VERSE was the agreed-upon name.
- There was some confusion regarding the general category of the character. Po was suggested with Terminal\_Punctuation=true, however, a similar character, U+06DE ARABIC START OF RUB EL HIZB, is So. The proposal should be given to PAG for consideration. Collation will depend on the character category.
- Mention that U+2055 FLOWER PUNCTUATION MARK or other various stars can be used to represent the star-shaped characters on page 1 that are also end of verse markers.
- Include the script identification. Currently the script identification should be Arabic, but if it is found with other scripts, then it can be changed to Common.

---

## 2 Sumero-Akkadian Cuneiform

**Action: For UTC discussion and decision**

**Document:** [L2/23-071](#) Unicode Cuneiform Sign Lists – Leroy

**Recommendations:** We recommend the UTC make the following disposition:  
We recommend the UTC to authorize the proposed draft UTR based on [L2/23-071](#).

**SAH-UTC175-R2:** Consensus: The UTC authorizes the Proposed Draft UTR Unicode Cuneiform Sign Lists, based on [L2/23-071](#). (Reference: Section 2 of L2/23-083)

**Comments:** We reviewed a proposal for new UTR: Unicode Cuneiform Sign Lists and are in favor of it.

The discussion centered on whether the proposed document should have the form of UTN or UTR. The document concerns an aspect of character identity and as such should go through UTC approval, hence as a UTR.

---

---

## II. SCRIPTS FROM EAST ASIA

### 3 Khitan Small Script

**Action: For UTC discussion and decision**

**Document:** [L2/23-065](#) Proposal to encode a blank character for Khitan Small Script (WG2 N5205) – West

**Recommendations:** We recommend the UTC make the following disposition:

**SAH-UTC175-R3:** Consensus: Provisionally assign U+18CFF KHITAN SMALL SCRIPT CHARACTER-18CFF, with name, annotation, and glyph as given in L2/23-065. (Reference: Section 3 of L2/23-083)

**Action Item** for Ken Whistler, UTC: Update the Pipeline to add U+18CFF KHITAN SMALL SCRIPT CHARACTER-18CFF as a provisionally assigned character. (Reference: Section 3 of L2/23-083)

**Comments:** We reviewed this request for a character that indicates a lost or illegible character. The glyph for the proposed character is a white box, similar in appearance to U+25A1 or U+25AF, and similar in purpose to U+13443, but has different properties and display behavior.

The SAH saw an earlier version and we were satisfied that the previous comments from the SAH had been addressed in this revised version (posted as L2/23-065).

For the glyph, Michel Suignard can use a generic square box glyph such as U+25A1.

---

## 4 Tangut

**Action: For UTC discussion and decision**

**Document:** [L2/23-066](#) Glyph Corrections for Eight Tangut Ideographs (WG2 N5206) – West

**Recommendations:** We recommend the UTC make the following disposition:

**SAH-UTC175-R4:** Consensus: Accept the glyph changes for 7 Tangut ideographs (U+17105, U+172A4, U+17BD1, U+17BD2, U+17BD3, U+17EF9, and U+18136) and update the Tangut data file for U+17105, based on document L2/23-066 for Unicode version 15.1. (Reference: Section 4 of L2/23-083)

**Action Item** for Michel Suignard, PAG: Update the Tangut data file for U+17105. (Reference: Section 4 of L2/23-083)

(No Action Item is needed for the font, as Michel Suignard already has received one from Andrew West.)

**Comments:** We reviewed this document proposing glyph corrections for 8 Tangut ideographs. This document has addressed comments the SAH made on an earlier version. Experts in China have agreed with the proposed glyph changes, except for U+18171. The Tangut data file needs to be changed because the new glyph for U+17105 has a different stroke count.

---

---

## III. SYMBOLS

### 5 Alchemical Symbols

**Action: For UTC discussion and decision**

**Document:** [L2/23-069](#) Revised designs of the alchemical symbols block – Miller

**Recommendations:** We recommend the UTC make the following disposition:

**SAH-UTC175-R5:** Consensus: Accept the glyph changes for the Alchemical Symbols block as given in L2/23-069 for the 15.1 version of the standard. (Reference: Section 5 of L2/23-083)

The following actions are recommended:

**Action Item** for Debbie Anderson, SAH: Send the font to Michel Suignard and the Charts group. (Reference: Section 5 of L2/23-083)

**Action Item** for Debbie Anderson, EDC: Prepare a glyph erratum notice for alchemical symbols for 15.1. (Reference: Section 5 of L2/23-083)

**Comments:** We reviewed this revised proposal to update glyphs in the Alchemical Symbols block. The proposed changes will bring consistency to the glyphs in the block: the glyphs for the original Alchemical Symbols, added in Unicode 6.0, do not match well with the new characters added in 15.0.

The version seen by the Script Ad Hoc addresses comments from the previous SAH meeting. The proposed glyph changes have been reviewed by experts.

The Unicode 15.0 glyph for U+1F763 ALCHEMICAL SYMBOL FOR PURIFY is a distinct glyph variant that is based on manuscript evidence. A concern was raised about U+1F763 being used in existing data as a variant of U+260B DESCENDING NODE. The SAH recognizes U+1F763 and U+260B as historical adaptations of the same underlying symbol. For U+1F763 in alchemical usage, a glyph shape similar to that shown for U+260B is more representative of actual practice. Users who desire the variant glyph can select fonts that reflect manuscript usage.

Because some of the proposed glyphs vary significantly from the glyphs in the current code chart, a glyph erratum notice would be useful.

---

## 6 Symbol: FORINT SIGN

### For UTC discussion and decision

**Document:** [L2/23-060](#) Proposal to Encode a Hungarian Forint Symbol in the Unicode Standard -- Nules, Vacek

**Recommendations:** We recommend the UTC discuss this topic during plenary or take this topic up later, after the SAH and PAG have met together.

**Comments:** We reviewed a proposal to encode ₦ as a currency symbol, which is similar to U+20A8 RUPEE SIGN ₨ and U+20A7 PESETA SIGN ₧.

We had previously declined to encode zł, a similar character that was proposed in L2/22-092, due to weak evidence. However, we felt that the evidence in L2/23-060 is stronger.

Since this is a legacy computing symbol, we propose to encode the symbol in the Symbols for Legacy Computing block instead of the Currency Symbols block, with properties as follows:

- U+1CCFA codepoint
- General category So
- Bidirectional class ON or L as per PAG recommendation
- Compatibility decomposition to Ft

An alternative approach would be to recommend use of ZWJ, as was recommended for a few Smalltalk characters described in section 4 of L2/21-234.

Depending upon the outcome of the discussion, it might be appropriate to ask the author if he finds any problem using ZWJ for round-tripping? What is the attestation of legacy implementations using the symbol as an atomic character? What is the communication process?

At the SAH meeting on April 7, 2023, we recommended the UTC provisionally assign U+1CCFA FORINT SIGN, based on L2/23-060 with properties as listed above. However, because the PAG members hold different views on how best to represent this character, we recommend discussion in the UTC plenary or in a separate PAG - SAH meeting.

---



---

## IV. PUBLIC REVIEW FEEDBACK

### 7 Arabic

**Action: For UTC discussion and decision**

**Documents:** Email Feedback (see text below)

[L2/23-061](#) Request to replace "start" with "star" in the name for U+06DE – Kamal Mansour

Date/Time: Mon Jan 23 11:25:12 CST 2023

Name: Michele Donini

Report Type: Error Report

Opt Subject: The Unicode Standard, Version 15.0

Unicode charcter U+06DE, named "ARABIC START OF RUB EL HIZB", should instead be named "ARABIC STAR OF RUB EL HIZB" (with Star, not Start).

Kind Regards,  
Michele D.

**Recommendations:** The SAH recommends to the UTC the following disposition:

**Action Item** for Lorna Evans and Kamal Mansour, EDC: Add text to Core Spec on U+06DE ARABIC START OF RUB EL HIZB in section 9.2, based on L2/23-061 and Section 7 of L2/23-083.

**Action Item** for Lorna Evans, Kamal Mansour, and Ken Whistler, EDC: Propose an annotation to U+06DE ARABIC START OF RUB EL HIZB in the names list for 15.1, based on L2/23-061 and Section 7 of L2/23-083.

**Comments:** We reviewed this feedback on the name for U+06DE ARABIC START OF RUB EL HIZB, a character that dates to Unicode 1.1.

While it appears at first glance that the name is an error for “STAR”, Mansour’s investigation discovered the name “START” is not an error, because the character indicates boundaries of parts or sections of the Quran and may appear at different intervals (quarters, halves, three quarters, etc.). The character can appear at the start or end of a section. The glyph itself is typically an 8-sided decorative star, but is not necessarily a star.

An annotation to U+06DE would be useful.

---

## 8 Bima

**Action: No Action Required**

**Document:** [L2/23-078](#) Comments on Public Review Issues (see text below)

*Related document:* [L2/23-070](#) Proposal to encode Bima characters – Febri Muhammad Nasrullah

Date/Time: Sun Mar 26 14:01:05 CDT 2023

ReportID: [ID20230326140105](#)

Name: Eduardo Marín Silva

Opt Subject: Feedback on Bima characters ([L2/23-070](#))

On document <https://www.unicode.org/L2/L2023/23070-bima-script.pdf> A few new characters are requested for the Buginese script to support the Bima orthography. I would just like to make a few suggestions.

1. Disunify the flower shaped end of section character: It is clear that this punctuation mark is shaped very differently than the existing end of section already encoded, with both signs having very different epigraphic derivations. It is quite likely that users won't see these characters as interchangeable and would like to use a particular glyph without switching font or even want to use them concurrently and/or outside a Bima context. Therefore they should be considered distinct. I recommend using the codepoint 1A1D, with the name BUGINESE FLEURON, or BUGINESE SIGN FLEURON to parallel the similar character 10AF1 MANICHAEAN PUNCTUATION FLEURON. A note could be added saying "used as an end of section sign in Bima texts". Along with the Pallawa and the existing end of section, they would be placed under the header "Punctuation". If the codepoint above needs to be occupied, I would recommend placing the

Reduplication sign or the Geminatio sign, since they are the most likely to see use beyond a Bima context. But it can also be left vacant for a future addition.

2. Disunify the killer above and below: While the author states that these are identical in meaning and use, the difference in placement may be very relevant for rendering the text. Not only epigraphers would like to represent the text as it was written, regardless of semantic distinction, it is generally very problematic for fonts and rendering engines to handle a combining mark that doesn't have a definite position with the base letter. It can result in glitchy rendering, particularly when it has to interact with other marks above or below. One could say that a vowel silencer would never be placed in the same letter as a vowel sign or a gemination sign, but if the Buginese script increases in popularity (as the author presumably wants it to), then more additions are likely, including ones that may be placed with the vowel silencer. The two glyph variants of the sign don't need to be disunified, they can be handled at the font level without issue; but the choice of one glyph over another in the code chart needs some justification.

3. Finally, the name "killer" needs to be reconsidered. While it is unlikely that people would take offense to it, it is suboptimal on its description. Better alternatives may be "Virama" or "Vowel Silencer".

4. Add a note under the Pallawa sign stating "a glyph variant with two dots occurs" or "may have three or two dots". These glyph variants are similar enough that they don't merit disunification.

5. Another modification would be to the one under 1A10, where instead of spelling it /h/, the note would read "used in Bima for ha".

In summary my most important suggestions are for the disunification of two more characters (the fleuron sign and the two positional versions of the vowel silencer), as well as adding a note under the Pallawa sign and reconsidering the name of "killer".

**Comments:** This feedback has been sent to the author of the Bima proposal [L2/23-070](#) and will be taken up again when the Bima proposal is considered by the SAH. (The proposal document L2/23-070 is currently posted for public comment, as noted below in Section 10.)

---

## 9 Sharada

**Action:** FYI with action to record

**Document:** [L2/23-078](#) Comments on Public Review Issues (see text below)

**Related document:** [L2/17-428](#) Proposal to encode the INVERTED CANDRABINDU for Sharada

Date/Time: Sun Mar 26 14:34:34 CDT 2023

ReportID: [ID20230326143434](#)



Name: Eduardo Marín Silva

Opt Subject: Request to revise the glyph for the character 111CF (SHARADA INVERTED CANDRABINDU)

On document [L2/17-428](#)

(<http://www.unicode.org/L2/L2017/17428-sharada-inv-candrabindu.pdf>) a proposal to add the Inverted Candrabindu for Sharada was made and eventually accepted. Unlike other Indic scripts, the regular Candrabindu for the Sharada script (11180) has an upper arc going above the dot, rather than a lower arc below the dot. In effect the Candrabindu is inverted with respect to the usual configuration. But it was brought to the attention of Unicode that a "regular" Candrabindu (inverted in the context of Sharada) was used along side the "inverted" (regular in the context of Sharada) and with different uses.

However the author decided to use a glyph with an arc that is shorter than its already existing dual. But an examination of the manuscript attestations in the proposal document, reveals that this isn't a consistent practice due to scribe idiosyncrasies: While figure 1, 2, 3 and six show the preferred glyph, figures 4 and 5 show an almost flat stroke instead of an arc, and figure 7 looks like an ark with an extra vertical stroke rather than a dot (contrasting with the regular one that does use a dot).

While the current glyph is fine, it makes one wonder if this discrepancy is important, and if making a font that just rotates the glyph of 11180 is acceptable. This situation is different than the Devanagari script, where the regular (0901) and the inverted version (0900) of the Candrabindu are mere rotations of each other.

I suggest contacting Anshuman Pandey and ask for his expert opinion, because if a glyph change is not warranted, an annotation noting the glyphic variants only applicable to the inverted version would be useful.

**Recommendations:** The SAH recommends to the UTC the following disposition:

**Action Item** for Debbie Anderson, SAH: Forward the SAH comments in Section 9 of L2/23-083 to the feedback author of Sharada feedback in L2/23-078 (Sun Mar 26 14:34:34 CDT 2023).

**Comments:** We reviewed the feedback on U+111CF SHARADA INVERTED CANDRABINDU, which is arguing for a glyph change. We encourage the feedback author to submit a formal proposal with the requested change. Such a proposal should explain how the current representative glyph is confusing and why the decision on the glyph shape should not be in the hands of the font creator.

This feedback has been forwarded to Anshuman Pandey, author of [L2/17-428](#). (Pandey has since reported offline that he does not find the change necessary.)

---

---

## B. RECOMMENDATIONS FOR 16.0

The SAH recommends the UTC approve all the characters and scripts that are “accepted for a future version” on the Pipeline (identified by yellow and green coloring) for publication in Unicode 16.0. The SAH has confirmed with EDC and PAG that the content for 16.0 (as described above) can be ready for a Unicode 16.0 release in September 2024. Charts including the new repertoire have already been developed.

**Recommendations:** The SAH recommends to the UTC the following disposition:

**SAH-UTC175-R6:** Consensus: Approve all characters and scripts accepted for a future version for Unicode version 16.0.

**Action Item** for Ken Whistler, UTC: Update the Pipeline, identifying all characters and scripts “accepted for a future version” on the Pipeline as accepted for Unicode version 16.0.

---

---

## C. DOCUMENTS NOT REQUIRING UTC ACTION (by script, in alphabetical order)

### 10 Bima

**Document:** [L2/23-070](#) Proposal to encode Bima characters – Febri Muhammad Nasrullah

*Related documents:*

[L2/22-150](#) Proposal to Encode Bima Characters – Febri Muhammad Nasrullah

(See list of other related documents on page 1 of L2/23-070)

**Comments:** We reviewed this updated proposal to encode 9 Bima characters. The SAH has seen many iterations of the proposal.

We discussed the proposed names of 3 non-letter characters. Based on names used in already encoded scripts, we recommend the following name changes:

16EA6 BUGINESE LETTER REDUPLICATION

16EA7 BUGINESE GEMINATION MARK

Further comments by Norbert Lindenberg:

- The grouping in section 3 is very useful in identifying some characters that definitely need encoding (green and yellow) or an annotation (blue), or definitely are identical to already encoded Buginese characters (white).
- The grouping gets fuzzy in the orange group: Two of them are proposed for encoding, while the others are not. While there’s good discussion of why some characters earlier proposed by Anshuman Pandey are not proposed here, overall it’s not clear how the line is drawn between the two characters proposed for encoding and the others that are not proposed.
- It seems the description of group 1 on page 3 should also include na and ma, bringing the total to 13 characters.

- The discussion of the glyph shape of the killer on page 4 needs to mention at the start that it's about the killer. The discussion might be better packaged as a separate subsection (as it was in L2/22-150).
- The sources of the character tables on page 4 should be named and added to the Bibliography.
- The table starting on page 4 and finishing on page 5 should be kept on one page.
- The Bugis-Makassar manuscripts shown in the table on page 5 should be added to the Bibliography
- The description of the group of characters in yellow cells on page 5 should get its own number within the list of such groups.
- In the second example on page 6, "This text is to be read as <Bima text> bebe", the reduplication sign should be replaced with the character U+1A05 BUGINESE LETTER BA to clarify how it is read.
- It would be useful to have information on line breaking as prose, using the questions given in the template for new script proposals as guidance. Such information, along with similar information for the Buginese writing system, could enable a decision on whether to adopt line breaking at orthographic syllable boundaries (L2/22-080).

The proposal is unclear on whether the present-day user community has been consulted. We recommend the proposal to be posted publicly before any decision on the proposal is taken.

---

## 11 Egyptian Hieroglyphs

**Document:** [L2/23-109](#) Draft encoding proposal for an extended Egyptian Hieroglyphs repertoire – Suignard, et al.

**Note:** The proposal document above was not available to the SAH to review, only the “principles” document (Appendix A) contained in L2/23-109.

**Comments:** We briefly reviewed a “principles” document that will be rolled into a preliminary proposal. The text on the encoding principles was based on the discussion of Egyptologists at the January 2023 meeting in Berlin, primarily documenting their decisions regarding which characters should be considered. “core” (i.e., recommended for use by Egyptologists and to be included in widely used fonts). We noted that the principles document is not targeting an encoding committee and suggested that the proposal should be revised to reflect guidelines used for encoding decisions.

A concern was raised about encoding characters that would become categorized as “do not use.” We recommend Michel Suignard and Debbie Anderson review Do Not Use text in the Core Spec so it aligns with the latest proposal.

Michel Suignard also reported on his progress re-drawing glyphs, as requested by the Egyptologists; this work has delayed preparation of a proposal. There will likely be between 3,000-4,000 characters proposed, but the full proposal will not be ready for SAH review before the April UTC. The goal is to get the characters into Unicode 16.0. The preliminary proposal can be on the April UTC agenda as informational; it is expected to appear at the July 2023 meeting for UTC action.

---

## 12 Jurchen

**Documents:** [L2/23-068](#) Towards an Encoding of the Jurchen Script (=WG2 N5207) - West

*Related documents:*

[L2/09-164](#) Proposal to Encode the Jurchen Characters (WG2 N3628) - China NB

[L2/09-352](#) Additional Information for the Proposal on Jurchen Characters (WG2 N3688) - China NB

[L2/21-049](#) A Supplementary Proposal to Encode the Jurchen Characters in UCS (WG2 N5131) - China NB

**Comments:** We briefly reviewed this draft document that will lead to a proposal for a historic script used to write the Jurchen language. The document proposes a new data file JurchenSources.txt that will accompany the future proposal, with fields described on page 96 (with an example).

The repertoire in the document includes 896 characters and 50 radicals, compared to the earlier proposals from China which totaled 1391 characters. The author proposes a character-based encoding model, compared to the glyph-based approach in the proposals from China.

Section 5.2 proposes a change in the Roadmap allocation for Jurchen (to U+18E00..U+1918F) and adding Jurchen Radicals to U+18DC0.. U+18DFF; this will require Khitan Ideographs to be moved. The SAH will review the requested re-allocation after WG2.

We noted that the document has been reviewed by experts and will be discussed during the upcoming June 2023 WG2 meeting. The SAH is awaiting the proposal document, which will be prepared after WG2, before recommending any further action.

---

## 13 Khmer

**Documents:** [L2/23-025](#) Khmer orthographic syllables - Lindenberg

[L2/22-290](#) Khmer Encoding Structure - Hosken et al.

**Comments:** We reviewed comments from SAH group members.

SAH recommends breaking L2/22-290 into smaller pieces, which are actionable by implementers.

Since the proposal is seeking to reduce the number of ways text can be encoded, concerns were raised about compatibility with existing documents. We would like to have a recommended way of encoding Khmer that keeps current documents working as they have been (similar to how the SHRII ligature in Tamil still works using TAMIL LETTER SA but is now recommended using TAMIL LETTER SHA).

The group also expressed a concern about including text with details on BNF complexity in the Core Spec and suggested this be an external document. Action Item 174-A58, assigned to Lawrence Wolf-Sonkin et al, covers this issue. James Kirby is invited to join.

Peter Constable observed that limitations in commonly-used input methods hinders solutions to encoding and usability challenges for Khmer and many other South and Southeast Asian scripts. Work in the CLDR Keyboard working group may offer future solutions that could help, hence it would be helpful to UTC and SAH to understand more about the status of that work.

---

## 14 Latin

### 14a Initial Teaching Alphabet

**Documents:** [L2/23-102](#) Unicode request for Initial Teaching Alphabet -- Miller

*Related Document:* [L2/22-286](#) Proposal to Encode Latin characters for Initial Teaching Alphabet – Manulov

**Comments:** We reviewed this proposal from Kirk Miller, based on his earlier document dated to 2020. Miller’s doc refers to a lightly revised version of L2/22-286 by Nikita Manulov. The proposal from Miller requests 13 Latin additions to support the Initial Teaching Alphabet, an alphabet used in teaching English-speaking users to read. ITA was developed in the 1960s. The proposal recommended certain unifications with existing characters.

(Note: Characters for Initial Teaching Alphabet initially appeared in [L2/08-428](#).)

The following points were raised during discussion:

- We recommend the use of IPA g U+0261 rather than U+0067.
  - Get feedback from Harald Tveiten based on his research.
- 

### 14b Slovene Metelko Alphabet

**Document:** [L2/23-074](#) Proposal to Encode Slovene Metelko Alphabet 2022 -- Manulov

**Comments:** We reviewed this proposal to add 21 Latin characters to represent Slovene Metelko text. This alphabet was used from 1825 until 1833.

The following comments were made:

- Provide a use-case for encoding these letters.
- Who is the user community?
- How big is the body of printed materials published in his orthography?

Debbie Anderson has forwarded the comments to the proposal author. The proposal has been posted in the document register should additional information be provided that builds a strong case for encoding the letters.

---

## 15 Seal

**Documents:** [L2/22-279](#) UCS Seal Script Source Mapping Data – Cook

WG2 N5209 Considerations concerning the Small Seal encoding initiative - Suignard

WG2 N5211 14-Column Seal Script Glyph comparison chart 2017-12-06 – Suzuki Toshiya, Richard Cook (see <https://www.unicode.org/wg2/WG2-curdoc.html> for link to doc with data; note it is 420MB)

**Comments:** The SAH had an oral report on the progress of Seal Script encoding from Michel Suignard, He pointed the Seal documentation and chart from Suzuki and Cook (N5211) and Suignard’s own “Considerations” document (N5209), all of which will be discussed at the June 2023 WG2 meeting. We welcome the progress that has been made, and encourage experts in Seal to work together to resolve outstanding issues. We look forward to seeing hope to see additional progress soon.

The SAH would like to stay informed, but relies on the expertise of others to evaluate any proposals regarding this script. Ken Lunde relayed offline that the CJK & Unihan Group wouldn't handle Seal script directly, but rather would encourage Seal experts to meet to review documents.

---

## 16 Sidetic

**Document:** [L2/23-019](#) Revised proposal to encode Sidetic in Unicode

**Comments:** PAG noted an error in the linebreak value on page 7. The value should be “AL” and not “R”.

This correction has been made by Anshuman Pandey and the corrected document has been uploaded into the document register: <https://www.unicode.org/L2/L2023/23019-sidetic.pdf>.

PAG has started an issue to create data for this script, which is only “provisionally assigned.”

---

## 17 Tulu

**Document:** [L2/23-055](#) Disunification of Tulu-Tigalari script & Invented Tulu Script – Vaishnavi Murthy, Vinodh Rajan

*Related documents:*

[L2/23-021](#) Additional Document for Unicode Tulu Proposal – Akash Raj Jain

[L2/23-012](#) Recommendations to UTC #174 January 2023 on Script Proposals – Anderson et al.

**Comments:** We discussed L2/23-055, which was a response from the Tulu-Tigalari authors to SAH Recommendations (pp. 7-8, L2/23-012) on how to proceed, since the UTC had already accepted Tulu-Tigalari and there is a request for modern Tulu (L2/23-021). Tulu-Tigalari is in CDAM2.

A few comments for the proposal authors of L2/23-055 on their doc:

- Explain the use of the triangles in the chart on page 5 (which indicate behavior mirroring Kannada or reformed Malayalam)
- Make it clear the Tulu-Tigalari script is clearly different from Tulu.

If the decision was made to disunify the scripts, one approach would be to re-name Tulu-Tigalari “Archaic Tulu-Tigalari” or “Old Tulu-Tigalari”, so there is no confusion with modern (reformed) Tulu. However, a ballot comment would need to be made on ISO ballots. To help make a decision on unification, examining one or two problems would be useful, such as how U is represented in the following:

Letter/Cluster	Modern form	Tulu-Tigalari
tu	ᱠᱤ	ᱠᱤ

However, the general feeling in the group was to retain the name Tulu-Tigalari as approved by the UTC (and as contained in CDAM2). If Tulu later is deemed stable, further discussion can take place on whether it should be unified or not.

In identifiers (IDNA), Tulu-Tigalari would have the Identifier type as Exclusion, as is typical for other historic scripts. If it were later to be unified with Tulu, the Identifier\_Type could be re-examined. (See [https://www.unicode.org/reports/tr39/#General\\_Security\\_Profile](https://www.unicode.org/reports/tr39/#General_Security_Profile) ).

Debbie Anderson responded to K TSA on Feb. 18, 2023, that the Script Ad Hoc would like to see more indication of stability in the script before recommending encoding (i.e., the proposed orthography is shown in publications by various authors and over a period of time). A list of printed books with the orthography as proposed, for example, would be important to include in a proposal.

---

## 18 Other topics:

### 18a Do Not Emit

**Comments:** This is an update to a proposed data file formalizing sequences of characters that should not be emitted by implementations, incorporating feedback from the SAH.

Feedback to the author Roozbeh Pournader included:

- document the possible values of Field 2 (with a key)

- add a header to the file
- address *eyelash RA* and canonical singletons
- discuss implications for other parts of the standard (data files, UAX #44 and UAX #31, UCD, etc.) (which will need to be reviewed by PAG).

Ned Holbrook will provide Roozbeh Pournader with language on emoji, basically directing people to UTS #51 instead of the DoNotEmit document.

If there is consensus in the UTC on DoNotEmit, the Core Spec text will need to be adjusted accordingly.

Roozbeh Pournader encourages feedback from UTC members.

---

## 18b Encoding Proposal Templates

### Documents:

[L2/23-104](#) Draft Template for character additions to an existing script in Unicode - Anderson

[L2/23-105](#) Draft Template for new script proposal - Anderson

**Comments:** We briefly reviewed the two draft templates for a new script and character additions. These can be used by new proposal authors in writing a proposal and can be linked from the [Submitting Character Proposals page](#). The SAH recommended they be posted in the document register and presented at the April UTC.

The UTC may wish to discuss whether a proposal author should identify their funding source in a proposal.

---

## 18c ScriptExtensions

**Comments:** The SAH discussed what the level of attestation is needed for ScriptExtensions, based on comments from Cibu Johnny and Asmus Freytag. Such information is useful for font designers and those working on script itemizers and rendering engines. In some cases, not enough information is known to be able to confidently list everything. For example, characters in the Combining Diacritical Marks block are not currently included.

Script Exemplars, which is a superset of orthographic exemplars in CLDR and vastly expands upon information in ScriptExtensions, is accessible from: <https://github.com/roozbehp/unicode-data/tree/master/exemplars>

After discussion the following general guidelines were agreed upon (captured by Manish Goregaokar):  
The characters in the ScriptExtensions file err on the side of being more inclusive rather than exclusive for scx, since it affects rendering engines (who may choose to show dotted circles /



brokenness when scripts are mixed). The bar here will be *attested* usage rather than *regular* usage, though CLDR may choose to be more conservative and stick to regular usage.

An interesting example that was brought up was Hebrew *niqqud* used in Arabic text and vice versa. Arabic harakat is also attested in Syriac.

Adding a header to the file describing its purpose and contents is recommended.

Mark Davis and Robin Leroy are invited to add in their comments.

---

---

## D. FYI DOCUMENTS

### Arabic

**Document:** [L2/23-064](#) Left-to-right directionality in Arabic numeric expressions – Kamal Mansour

**Comments:** This document answers a question posed in Part 2 of L2/22-261. It has to do with problems handling certain text in bidi contexts, specifically sequences that involve combinations of digits and punctuation that do not represent numeric values, such as phone numbers. Debbie Anderson has forwarded this document to the author of L2/22-261, and will also bring it to the attention of the CLDR chairs as of possible interest for CLDR.

---

### Badugu

**Comments:** We discussed a draft proposal for Badugu, a script that was first introduced in L2/22-220. Subsequent information has been received that raised questions whether the proposed script is in widespread usage and if there is an independent user community. More evidence would be needed before proceeding, including confirming this script is the preferred one by the community. It was noted that there are two other scripts (though the proposal only mentions one).

---

### Old Hungarian

**Document:** [L2/23-067](#) Article on Old Hungarian letters for Q, X, Y -- Viktor Kovács

**Comments:** This is an FYI document to the UTC. It addresses the earlier request to encode Old Hungarian Q, X, and Y, which was recommended in L2/21-115.

## E. IN PROCESS

### Other script and character topics in process

The following script and character topics are in process:

- Incung