

Release Management Group Report to UTC #181

November 6, 2024

Peter Constable & UTC Release Management Group

The Release Management Group (“RMG”) is composed of UTC working group leaders and other volunteers working on the process for development of the Unicode 17.0 release and for evolving processes for future Unicode Standard releases.

Unicode 16.0 release

Unicode 16.0 was released on September 10, according to plan. French charts are normally completed some time following the main release; the French charts for 16.0 have not yet been published.

Process development

During a retrospective exercise after the Unicode 16.0 release, RMG reviewed what process improvements had been made during the Unicode 16.0 release cycle. The following had been identified last year as goals for the 16.0 cycle:

- CJK & Unihan: Ensure backup resources. (Currently, Ken Lunde alone has expertise and is a single point of failure.)
- Improve processes for release collateral

Significant progress was made on the first item: thanks to Ken Lunde, all of the tools and source data for the Unihan database are in the private unihan-tools GitHub repo, along with detailed documentation. Note that, at present, Ken is the only one that has actually performed these tasks.

Some progress was made on release processes, including some streamlining of the process for deploying UCD and other data files. Further improvements should be made.

The post-15.1 retrospective also considered processes and data management for code charts. This has been much improved: thanks to work by Liang Hai and the Charts WG, all of the fonts and snapshots of code chart source files are in GitHub repos. In addition, thanks to work by Liang and Asmus Freytag, tooling and processes are in place to enable representative glyphs from the charts for any character to be used in core spec content.

The post-15.1 retrospective also considered the need to contain the scope of releases—to slow down—so that we don't over-burden our volunteer contributors and put quality of releases at risk. On reviewing the 16.0 release, RMG noted that we did *not* slow down. One specific aspect was that the production work required for Egyptian Hieroglyph additions was much bigger than anticipated, particularly in regard to names list content.

When the 16.0 cycle began, drafts for much of the new content for the core spec had already been prepared. As we begin the 17.0 cycle, we are far behind where we were for 16.0 at the start of that cycle. While detailed block descriptions for new scripts are not generally considered blocking for a release (a short stub description can be sufficient), not providing a good block description in the same release as when a script is encoded only adds to the backlog of core spec work that should be done.

Noting these last two points, RMG strongly advises UTC to consider how best to manage the production work required for new repertoire, particularly in regard to content for the names list and for the core spec.

RMG recommends that, going forward, more burden for draft names list and core spec content be pushed upstream to proposal authors and the Script Encoding WG, that progress on this content be evaluated during the alpha and beta review periods, and that UTC be willing to postpone new character repertoire if this content is left to be completed late in the release cycle (any time after the start of the beta review period).

In the post-16.0 retrospective, RMG also reviewed a white paper prepared by Ken Whistler in September 2021 for Unicode leadership (following discussion of topics at UTC #168) regarding distribution of production workload and risks if we don't find new people to share the workload before certain people (especially Ken, Rick and Michel) retire. Significant progress has been made over the past three years, particularly with the formation of RMG and the Charts WG. Two of the areas that Ken had called out in 2021 are still a concern:

- “Technical report czar”: We have a growing number of UAXes and UTSeS that need to be maintained and updated for each release. While different working groups are responsible for most of these, there should be some centralized product management and consolidation of certain editorial or production tasks.
- Core spec managing editor: The core spec is a very large document with myriads of details that need to be kept coherent, and in any release there can be a large number of planned revisions to be tracked. In the past, this was handled by a paid staff person, not a volunteer.

These gaps warrant careful consideration by UTC and by Unicode officers.

Finally, in the post-16.0 retrospective, RMG considered what significant process improvement goals should be set for the 17.0 cycle. The following were identified:

- Refine the release deployment process (continuation of a 16.0 process goal). Some preliminary ideas include:
 - More automation of deployment actions.
 - Staging of all release collateral well in advance of the release to facilitate validation, reduce day-of-release tasks
- Continue to expand shared understanding of all “big red switch” production tasks and distribution of task ownership so there is less dependency on Ken Whistler.

As part of the transition from the 16.0 cycle to the start of the 17.0 cycle, RMG reviewed all BRS tasks assigned to Rick McGowan with a view to re-assigning them to new owners. That has mostly been completed; there are a small number of tasks that the Editorial Working Group still needs to review.

Unicode 17.0 scope and timeline

Unicode 17.0 should be planned as a “full” release including new characters and scripts. This will already be a large release, in terms of new/revised content to be prepared. A backlog of provisionally-assigned scripts and other characters has been accumulating, and we should include a large number of these to reduce this backlog. At the same time, changes to many aspects of the development process have been on-going. Also, we should maintain the annual September release schedule. Thus, we need to balance the scope between reducing the backlog while ensuring we can have high confidence of completion on schedule. Hence RMG advises UTC to be conservative in adding any additional new repertoire beyond that proposed in this report (see below).

A timeline for the development and release of Unicode 17.0 is proposed with the following key dates:

- 2024-11-8, UTC #181: Confirm scope and timeline for Unicode 17.0
- 2025-1-10: snapshot of PRI feedback
- 2025-1-24, UTC #182: Finalize alpha content
 - New characters / emoji, charts
 - Property data
 - Advance in-progress UTSeS, UAXeS
- 2025-2-4: Start of alpha
- 2025-4-1: End of alpha, snapshot of feedback
- 2025-4-24, UTC #183: Finalize beta content
 - Charts
 - Property data
 - Draft UTSeS, UAXeS
 - Draft of complete core spec content
- 2025-5-20: start of beta review
- 2025-7-1: end of beta public review

- 2025-7-24, UTC #184: Finalize 17.0 content
- 2025-9-9: Unicode 17.0 release

RMG proposes that UTC #181 take the following action:

[181-C] Consensus: UTC approves the proposed timeline and plan for Unicode 17.0 outlined in document L2/24-258.

Recommendation of new repertoire for 17.0

In RMG's report to UTC #179 (L2/24-070), we gave a preliminary recommendation for new repertoire to be encoded in Unicode 17.0. Subsequently, UTC #180 took the following actions accepting characters for version 17.0:

[180-C1] Consensus: Accept 4,300 CJK Unified Ideographs in a new CJK Unified Ideographs Extension J block with code point range U+323B0..U+3347B (block: U+323B0..U+3347F), based on document L2/24-138 (aka WG2 N5254) and Section 07 of document L2/24-165, for Unicode Version 17.0.

[180-C2] Consensus: Accept the disunification of U+5CC0, the encoding of its disunified form at U+2B73A at the end of the Extension C block, and all of the associated property value changes and additions, based on document L2/24-138 (aka WG2 N5254) and Section 07 of document L2/24-165, for Unicode Version 17.0.

Thus, 4,301 characters are already accepted for encoding in Unicode 17.0, but none of the characters previously given provisionally assigned code points are as yet accepted for Unicode 17.0. RMG recommends that UTC approve 507 characters for encoding in Unicode 17.0 for which code points have been provisionally assigned, as described in the proposed actions shown below.

Note that ISO/IEC JTC1/SC2 has circulated a draft of a new edition of 10646 for comments, which includes all of the above characters plus those approved at UTC 180. National body comments could lead to additions or deletions in that draft. Additions would not necessarily affect UTC plans for Unicode 17.0 since the new edition of 10646 is planned to synchronize with Unicode 18.0. If national bodies request characters be removed, however, we may want to remove some of the proposed characters from 17.0. RMG has taken this into consideration when preparing this recommendation.

RMG proposes that UTC #181 take the following actions:

[181-C] Consensus: UTC accepts for encoding in Unicode 17.0 the following 228 characters for new scripts for which code points have previously been provisionally assigned: 50 characters for Beria Erfe script, 40 characters for Chisoi script, 29

characters for Sidetic script, 55 characters for Tai Yo script, and 54 characters for Tulong Siki script.

[181-C] Consensus: UTC accepts for encoding in Unicode 17.0 the following 283 Arabic, Bengali, Han, Kannada, Latin, Oriya, phonetic, Sharada, Tangut, Telugu, and other symbol characters for which the code points have previously been provisionally assigned:

- Arabic (40 characters): 088F, FBC3..FBD2, FD90..FD91, FDC8..FDCE, 10EC5..10EC7, 10ED0..10ED8, 10EFA..10EFB
- Bengali (1 character): 09FF
- Han (5 characters): 16FF2..16FF6
- Kannada (1 character): 0CDC
- Latin (2 characters): A7D2, A7D4
- Oriya (2 characters): 0B53..0B54
- Phonetic (18 characters): 1ACF, 1AD0..1ADD, A7CE..A7CF, A7F1
- Sharada (8 characters): 11B60..11B67
- Tangut (143 characters): 187F8..187FF, 18D09..18D1C, 18D80..18DF2
- Telugu (1 character): 0C5C
- Symbols (62 characters): 2B96, 1CCFA..1CCFC, 1CEBA..1CED0, 1CEE0..1CEF0, 1F777..1F77A, 1F8D0..1F8D8, 1FA54..1FA57, 1FBFA

[181-A] Action item for Ken Whistler, RMG: Update the pipeline to reflect the assignment of 511 provisionally assigned code points for encoding in Unicode 17.0. [Ref. consensus 181-C??, 181-C??]

[181-A] Action Item for Markus Scherer, PAG: Update the relevant UCD data files in preparation for Unicode version 17.0 to include 511 additional characters approved for encoding. [Ref. consensus 181-C??, 181-C??]

Note: RMG anticipates that, during UTC #181, the CJK and Emoji working groups could independently propose that additional characters be accepted for encoding in Unicode 17.0.