

# Release Management Group Report to UTC #174

January 19, 2023

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 15.1 release and for evolving processes for future Unicode Standard releases.

## Unicode 15.1 timeline

A timeline for the development and release of Unicode 15.1 was proposed in [L2/22-270](#) and confirmed by UTC #173 (cf. [173-C1](#)). The following is a summary of key dates:

- 2023-1-26, UTC #174: Finalize alpha content
- 2023-2-7: start of alpha review
- 2023-4-27, UTC #175: Finalize beta content
- 2023-5-23: start of beta review
- 2023-7-27: UTC #176: Finalize 15.1 content
- 2023-9-12: Unicode 15.1 release

Planning discussed by UTC #173 considered releases for both 2023 and 2024, with a scaled back 15.1 release in 2023 anticipating the need for changes in volunteer roles and adjustments over these two years. To help communicate awareness to implementers and stakeholders, a Unicode blog post was published in mid-November: [The Unicode Blog: The Unicode® Standard – 2023 Release Planning](#).

## Unicode 15.1 alpha review

The 15.1 alpha review period will begin soon after UTC #174. The RMG has begun preparing for this milestone. Key deliverables are expected to consist of (i) code charts, (ii) a subset of data files (UCD, Emoji), (iii) a PRI cover page, and (iv) a Unicode Blog post.

Preparation of the alpha content is dependent on the UTC formally deciding on the content to be included in the alpha. The RMG assumes that, as in the past, this will minimally consist of the core UCD data files (which establish the repertoire), and draft delta charts for new repertoire or glyph changes to existing repertoire. Based on input from Markus Scherer regarding PAG progress, we recommend that the alpha include core UCD files but not other groups of data files

— see [L2/23-039](#). Similarly, based on work within ESC, we recommend including draft Emoji data files — see [L2/23-010](#).

## Progression of new script and character repertoire

The RMG recommends that UTC adopt a slightly revised process for how new script and character repertoire advance from encoding proposal to publication in a new Unicode version. This is intended to provide a better balance of commitments with available resources going forward without necessarily introducing any delay.

The usual process in the past has been:<sup>1</sup>

- 1) Encoding proposals are vetted by SAH.
- 2) When mature, SAH recommends UTC to *accept characters for encoding*.
- 3) UTC accepts characters for encoding in the next version under development, or in a “future” version (typically understood to be the following version).
- 4) Characters accepted by UTC immediately enter formal processes in ISO.

The revised process we recommend would be as follows:

1. Encoding proposals are vetted by SAH.
2. When mature, SAH recommends UTC to *reserve code points*. SAH also notifies PAG to begin review of properties.
3. Preparation of collateral needed for publication — code charts, property data, block descriptions or other core text changes — is done based on reserved code points.
4. When collateral is sufficiently mature, UTC accepts characters for encoding in a specific version.
5. Characters enter formal processes in ISO.

*Note: This proposal was brought to SAH at a recent meeting, and SAH recommendations for UTC #174 ([L2/23-012](#)) were prepared assuming that UTC would adopt this recommendation.*

In practical terms, there would not be a large difference from how work currently progresses on the Unicode side. We would, however, be more explicit in choosing target releases, including what ISO ship vehicle we would like to see targeted.

---

<sup>1</sup> In the past, a common variation of this process was for a proposal to get vetted first within JTC1/SC2/WG2, then recommended for addition to the next amendment or edition of ISO/IEC 10646, and then introduced to UTC in a “consent docket” to keep Unicode and 10646 in sync. Apart from CJK, this has not happened in the past few years.

To understand the reasons behind this recommendation, it will be useful to consider how things were in the past and how our context has been changing.

Historically, repertoire additions in Unicode and ISO/IEC 10646 were closely synchronized and developed in parallel. In the past, both UTC and SC2/WG2 met on regular and frequent schedules, and there were usually plenty of volunteers contributing time to prepare all the content needed for publication—charts, data files, block descriptions, etc. In this paradigm, new script and character additions advanced through formal steps on both Unicode and ISO sides in parallel. Before new repertoire were published in a new Unicode version, it would always have gone through an Enquiry Stage (“DAM” or “DIS”) ballot in ISO. Similarly, any new repertoire introduced in WG2 and added to an ISO draft amendment would get added to Unicode’s process, and our volunteers worked diligently to make sure everything was ready for the next release.

Our context has evolved in recent years, however. COVID-19 was certainly a factor: the pandemic made it impractical to conduct WG2 meetings; and even apart from meetings, activity within WG2 has slowed.<sup>2</sup> Another significant factor has been changes in volunteer resources that, in the past, got the bulk of the work done—e.g., Ken Whistler stepping back from certain roles. We have needed to learn that we cannot take our volunteer resources for granted and cannot assume there will always be someone getting everything done. Our changing context brings less certainty about capacity to get all content ready in a short time.

Given this, we need to be more thoughtful about the repertoire we commit for specific Unicode versions and ISO ship vehicles. On the one hand, there is less content that needs to be prepared for ISO 10646 than for Unicode—no block descriptions, and much less in the way of data files. But on the other hand, inserting new repertoire into an ISO amendment imposes some constraints: ISO amendments are required to progress on a certain timetable, and once the characters are included into draft ISO charts it becomes difficult (given existing processes) to shift them to a different target.

Longer-term, the RMG envisions future processes in which the various types of content needed for publication (charts, data files, core spec text...) can be developed as volunteer resources are available with agility regarding the release vehicles until all the content is sufficiently ready to commit to target releases of Unicode and ISO 10646. We are not yet at that point in our processes, but the RMG considers the recommended change in formal UTC actions will be a helpful step at this time.

---

<sup>2</sup> This is evident by looking at the [WG2 Document Register](#) and the rate at which documents have been submitted over recent years.

# Evolving development processes for Unicode 15.1 and 16.0

RMG, along with other UTC work groups, have looked for opportunities to make modest process changes that could provide greater agility without putting progress for the 15.1 release at risk. These are some of the changes that have been made:

- “Big red switch” task list: This is a large list of tasks involved in each phase of development for each Unicode release. In the past, this was maintained by one person (Ken Whistler) in an HTML page on the unicode server. This has been migrated to a Google sheet to allow easier, collaborative maintenance.
- Like other UTC work groups, RMG has started using GitHub issues to manage tasks not directly tied to the release.
- Processes for UAXes: In the past, EdComm drove development for all of the UAXes as well as for UTSes and UTRs synchronized with the release. Ken Whistler would track the status of each doc in an Excel file maintained on the Unicode server. RMG took over overall responsibility for driving production of these docs, allowing EdComm to limit itself to editorial review (its mandate). RMG has started to push more responsibility for development tasks to the UTC work groups (PAG, ESC, CJK) that are the appropriate technical owners of these docs. RMG collectively has been tracking status of all the docs, maintaining the status info in a Google sheet.
- Code charts: Up to now, Michel Suignard and Rick McGowan have been the code chart “tsars”, with Ken Whistler helping to maintain the NamesList.txt file used by the chart tool set. A Charts work group has been formed to bring in additional people. The initial goal is to have others learn the chart development and production processes, and to ensure that these processes are well documented. As longer-term goals, the group will look at adapting tools and processes for additional purposes, such as providing assets needed for the core spec.
- Data files: For many years, the data files were developed primarily by Mark Davis and Ken Whistler, and various manual steps were involved in publishing the data files on unicode.org. In recent years, Markus Scherer and others have taken on more of the process and maintenance of tools used for development of the data files, though the same manual steps have been required for publishing of the files. A tools team under PAG has continued to work on tools and streamlining of processes. Work is being done for 15.1 to bring more automation into the publishing process.

Evolution of our processes is on-going.

[End of report]