

Title: Proposal to normatively reference external code charts
Status: Follow up on recommendation M72.10 (Code Charts)
Source Michel Suignard (ISO/IEC 10646 Project Editor)
Actions For consideration by SC2, WG2
Date: 2026-06-07

Summary

This document describes a process by which ISO/IEC 10646 could normatively reference a set of code charts produced by an outside entity (the Unicode Consortium) which would still be an integral part of the standard. This would simplify greatly the production of the standard by making it editable by a regular text processor. This can be achieved by adding a normative reference to a new resource (Code chart) and making sure that the reference matches the repertoire covered by this edition of ISO/IEC 10646 and its amendment(s) as applicable. This documents describes in detail the required modification corresponding to the 7th edition of ISO/IEC 10646 if the proposal is accepted.

Details

Modification to clause 2: Normative references.

In the context of Unicode 18.0 and ISO/IEC 10646 7th edition, the proposal is to recreate a parallel folder structure relying on the existing structure that was proposed in the past:

<https://www.unicode.org/wg2/iso10646/edition7/charts/>

(the folder content would be created in a manner similar to Unicode version with a block-by-block view and a consolidated view, but no delta view). (See discussion about code chart types below)

A new normative reference would be added to reference the code charts, such as:

ISO/IEC 10646 7th Edition, Code charts <https://www.unicode.org/wg2/iso10646/edition7/charts/>

Still open for debate is whether we refer to the Unicode 18.0 code charts directly, or whether we create a parallel structure with same repertoire content. While the repertoire is identical, it may be desirable to use different header or footer for ISO/IEC 10646. In addition, it would now be possible to create a block-by-block view of the standard, similar to what is available at <https://www.unicode.org/charts/>. Note that there is no intent at this point to create an index view similar to the link above but simply a folder containing all these blocks. Managing and maintaining a separate index view specific to ISO/IEC 10646 would be a complex endeavor. Note that ISO/IEC 10646 can still refer informatively to the Unicode index view as it addresses the same repertoire.

The Unicode Standard maintains 3 types of code charts for each version:

- Code charts (formerly un-versioned). These are block-by-block code charts of the whole repertoire and are typically accessed using an index page.
- Consolidated view (formerly archived) which contains in a single PDF file the whole repertoire. The file is bookmarked to allow easy navigation to individual sections of the PDF file. That PDF file is extremely large (approximately 180MB).
- Delta charts (formerly versioned) which are also block-by-block files. These blocks contain newly encoded characters or existing characters with significant changes

to existing glyphs or properties associated to the related code points (such as standardized variation sequences). These delta charts are also typically accessed by a single delta code charts page which points to the various blocks. Unlike the other code chart types, the blocks are presented with highlights (yellow for new character and blue for other changes).

Until now, ISO/IEC 10646 editions have only included code charts equivalent to the consolidated view but integrated in the standard itself which makes the standard size extremely large. Amendments have used a format not in use by the Unicode standard which is a simple PDF including all blocks of the Delta charts with the same highlights.

Modification to clause 14: Structure of the code charts and lists.

Replace clause title with “**Structure of the code charts**” and replace the first paragraph with the following new text:

The code charts, including the lists of character names, shall be in accordance with the referenced ‘ISO/IEC 10646 7th Edition, Code charts’ (see clause 2). They specify for each character its graphic symbol, its coded representation, and its character name, unless the name is defined algorithmically (see sub-clause 7.4).

Modification to clause 37: Code charts and lists of character names.

Replace clause title with “**Code charts**” and replace sub-clause 37.1, 37.2, and 37.3 with the following text:

37.1 General

The detailed code charts for the BMP, SMP, SIP, TIP, and SSP are shown in the referenced ‘ISO/IEC 10646 7th Edition, Code charts’ (see clause 2). They specify for each character its graphic symbol, its coded representation, and its character name, unless the name is defined algorithmically (see sub-clause 7.4). These code charts are available either as a single consolidated PDF file (consolidated.pdf) containing all blocks, or as a set of block-by-blocks PDF files in the form Ux.pdf, where ‘x’ represents the first code point in hexadecimal notation of the block. For example, the file name U4E00.pdf corresponds to the CJK Unified Ideographs block (4E00..9FFF).

There are three types of code charts depending on the types of character names, specified algorithmically or not, and the type of information provided in the code charts. These types are described in 37.2, 37.3 and 37.4.

37.2 Code chart for characters with names not algorithmically defined

Code charts are presented in block-by-block arrays of graphic symbols representing the characters organized in one to sixteen columns of sixteen graphic symbols each. The lower digit of the coded representation is indicated in the left margin while the remaining upper digits are indicated in the top margin. The full coded representation for each character is also indicated under each representative graphic symbols.

Note 1 Graphic symbols corresponding to the representation of graphic characters are informative. See Clause 14.

Note 2 Characters for which one or more standardized variation sequences are defined are indicated with a triangular mark in the top-right of the corresponding code chart cell.

(remove 37.3 Character names list sub-clause title)

These arrays of graphic symbols are followed by the character names list and a summary of standardized variation sequences (see 17.6.2) as appropriate.

Note 3 A array and related names list may be arranged in a single page if their contents allow it.

(follows content of current sub-clause 37.3, not repeated here)

(remove sub-clause 37.4 and 37.5 and create new sub-clause 37.3 and 37.4)

37.3 Code chart for characters with names algorithmically defined and no annotation

Most of the code charts with names algorithmically defined have no additional annotation.

Therefore, their code charts do not show their names as a separate names list. This covers the CJK Unified ideographs, the Tangut characters, the Jurchen characters, the Small Seal characters, and the Hangul Syllables. Their presentation is specified in 24.3 for CJK Unified ideographs, in 25.3 for Tangut ideographs, in 26.3 for Jurchen characters, and in 27.3 for Small Seal characters.

37.4 Code chart for characters with names algorithmically defined and annotation

Some code charts with names algorithmically defined have additional annotations and are therefore displayed differently.. The code chart for CJK Compatibility ideographs uses an hybrid model specified in 24.4, mixing the presentation used for CJK Unified ideographs, with additional annotation. Other blocks with names algorithmically defined, such as the Nüshu characters, the Khitan Small Script characters, and the Egyptian hieroglyphs for the extended block, use the same presentation as blocks using non algorithmically defined names specified in 37.2.

It should also be noted that ISO/IEC 10646 contains Annex G ‘Alphabetically sorted list of character names’ which is repertoire related but has no equivalent in the Unicode Standard. No action is required at this point.

Background context

In June 2025, SC2 in resolution M30-04 adopted the WG2 recommendation M72.10 which had the following text:

Recommendation M72.10 (Code Charts)

WG2 recommends that the Project Editor explore with ITTF the possibility of moving the UCS code charts from the core text of the CD into an electronic data attachment and, if an acceptable plan can be worked out, to implement that in the CD.

Following that, I communicated with ITTF (Mr. Alexander ESTRADA, ISO Editorial Manager), mentioning:

...I have done experiment of that (but not in the current CD (Cd.3)) and now with the synchronization between ISO/IEC 10646 and Unicode repertoire (for example the 7th edition content will be the same as Unicode v18.0), it would be extremely beneficial to share that code chart content between Unicode and 10646. The ISO/IEC 10646 code chart is in fact identical to the 'archived' version of Unicode of the synchronized version (for Unicode 17.0 it is <https://www.unicode.org/Public/17.0.0/charts/CodeCharts.pdf>). (Unicode v17.0 has no 10646 equivalent, Unicode v16.0 was synchronized with amd2 of the previous edition). Note that the file is protected and bookmarked by blocks which makes it much more accessible.

My preferred option would be to make that code chart pdf file a normative reference (clause 2 of the standard). The other option is to make it an electronic data attachments (clause 5). I prefer the clause 2 route because it makes the management of copyright issues simpler. For all practical purpose, the code chart development is entirely funded and created by the Unicode Consortium which has provided terms and conditions to provide adequate referencing by ISO, see <https://www.unicode.org/charts/About.html>, especially the sections 'Charts and Ancillary Data for ISO/IEC 10646' and 'Terms of Use' which says: 'The Unicode Consortium specifically grants ISO a license to produce such code charts with their associated character names list to show the repertoire of characters for that standard, as a normatively referenced, integral part of that standard.'

It would greatly simplify the editing work on this project as the code charts production is not even managed by Word, and would be even less manageable by OSD. While moving the rest of the standard to OSD is challenging in itself (due to its multi-lingual nature), at least it becomes an option.

The answer from Alexander ESTRADA was the following:

I concur with your assessment and agree with the Clause 2 route. I see similar maneuvers were employed in previous versions of ISO/IEC 10646. As with any Clause 2 reference, please ensure that it is cited by using a "shall" or "shall be in accordance with" statement, which denotes its use is a requirement of the document.

...

If the document is to be referred in Clause 2, please follow the [ISO/IEC Directives Part 2, 10.2 Permitted referenced documents](#) "Normatively referenced documents shall be documents published by ISO or IEC. In the absence of appropriate ISO or IEC documents, those published by other bodies may be listed as normative references provided that

- a) the referenced document is recognized by the committee as having wide acceptance and authoritative status,*
- b) the committee has the agreement of the authors or publishers (where known) of the referenced document for its inclusion as a reference,*

- c) the authors or publishers (where known) have also agreed to inform the committee of their intention to revise the referenced document and what points the revision will affect,
- d) the document is available under commercial terms which are fair, reasonable and non-discriminatory, and
- e) any patented item required for the implementation of the document is available to be licensed in accordance with the ISO/IEC Directives, Part 1, 2021, 2.14, "Reference to patented items".

...

The ISO Technical Program Manager attached to JTC1/SC2 (Mr. Stephane SAUVAGE), concurred by saying: "Route for clause 2 is indeed best to use to make normative reference to avoid duplication of effort..."

To which I answered:

To give some context, the code charts for 7th edition are eventually going to match the 'archived version of Unicode V18.0 code chart, version which is currently in its own 'alpha' release. The content of this code chart is currently available in draft form at <https://www.unicode.org/Public/draft/charts/CodeCharts.pdf> and will eventually show in <https://www.unicode.org/Public/18.0.0/charts/CodeCharts.pdf> when finalized. The official version of that archived version of the code charts for Unicode v17.0 is available at <https://www.unicode.org/Public/17.0.0/charts/CodeCharts.pdf>. (there is no ISO/IEC 10646 equivalent to Unicode V17.0)

It is my understanding that the linked document above is acceptable as a referenced resource per the [ISO/IEC Directives Part 2, 10.2 Permitted referenced documents](#). You can look at the first page of that file as it contains some terms concerning usage by ISO (this could be modified if needed). Because going from an electronic attachment to a linked reference goes beyond what was suggested in the WG2 recommendation and related SC2 resolution, this probably means that this solution cannot be implemented right away, but at least it can be discussed at the next SC2 plenary next June. This obviously implies that to be referenceable, the repertoire content needs to be 100% in sync between a Unicode version and an edition (or amended) of ISO/IEC 10646.

Then Stéphane answered:

"...

I believe on my side that the reference(s) to the Unicode Standard(s) can be listed in clause 2 as it is currently in the text of [ISO Projects - ISO/IEC CD 10646.3 - Overview](#), but they need to be mentioned in normative manner in the content of the text ...

Note that as indicated in 10.2 of directives part 2, the committee needs to have at hand the agreement of UNICODE, see b) and c) of 10.2 of directives part 2, extract below :

- b) the committee has the agreement of the authors or publishers (where known) of the referenced document for its inclusion as a reference,
- c) the authors or publishers (where known) have also agreed to inform the committee of their intention to revise the referenced document and what points the revision will affect,

...

Concerning the agreement, note that all Unicode code Charts contain the following text in their header:

The Unicode Consortium specifically grants ISO a license to produce such code charts with their associated character names list to show the repertoire of characters for that

standard, as a normatively referenced, integral part of that standard.

This should cover the item b) mentioned above. In the context of the Unicode standard and ISO/IEC 10646, the second item is covered by the repertoire synchronization between the two standards.

EOF