

Title: Considerations concerning the publication of ISO/IEC 10646 5th Edition

Source: Michel Suignard, Project Editor

Status: Individual Contribution

Distribution: WG2

References: SC2 N4506 WG2 N4777

Summary: This document is requesting an endorsement from SC2 to publish ISO/IEC 10646 under a new format as requested by ISO/IEC new directives.

One may be surprised why the 5th Edition of ISO/IEC 10646 has not been published yet. After all, the DIS was approved in November 2016, the changes requested were so minor that a FDIS ballot was unnecessary. The disposition of comments was published November 14 2016 (document WG2 N4777), more than 10 months ago.

However, when the publication process was initiated, the ISO publishing branch (ITTF) requested that the standard be published as a single PDF with regular hyperlinks pointing to ancillary data. In previous versions, ISO/IEC 10646 was published with ancillary data pointed by file link to files located in the same folder/directory as the main body of the text. Some of the files contain script specific data, others contain collection data, or information better represented as mechanically retrievable. These files are shown in the list below:

- UCSVariants.txt,
- EmojiSrc.txt,
- CJKSrc.txt,
- TangutSrc.txt,
- NushuSrc.txt,
- NUSI.txt,
- JIS-X-0213-FromPrevious.txt,
- JMJKI-2016.txt,
- Allnames.txt,
- HangulSy.txt.

In addition, given their size, the code charts were also accessed through multiple files. For the 5th edition, it had been decided to simplify the schema by just using a single large file:

- 10646-5th-01-CodeCharts.pdf (because the standard was downloaded as a single ZIP, there were no benefit in keeping the segmentation in multiple files for the code charts).

The new ITTF requirements meant that the data files had now to be normatively referenced from clause 2 (Normative references) as outside hyperlink (http), and that the code chart would also become an outside resource. In addition, the new ISO/IEC directives created new text requirements for Clause 2 (Normative References) and 3 (Terms and Definition).

Based on this, all these files had to be posted on a stable repository. We could either use an ISO based address with the usual ISO based restrictions or another external site. After some discussion with ISO staff it was determined that the optimal solution was to post these files on the Unicode web site with the understanding that the Unicode Consortium would follow the rules expressed in the ISO directives part 2 clause 10.2 (repeated below):

10.2 Permitted referenced documents

In principle, normatively referenced documents shall be documents published by ISO and/or IEC. In the absence of appropriate ISO and/or IEC documents, those published by other bodies may be referred to in a normative manner 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 and*
- d) the document is available under commercial terms which are fair, reasonable and nondiscriminatory.*

Informative reference may be made to any other type of document. Informative references shall be listed in the bibliography.

ISO and IEC normatively referenced documents shall have reached at least the enquiry stage (40.20 DIS or CDV).

The committees shall validate all referenced documents when a document is revised.

The normative references list shall not include the following:

- referenced documents which are not publicly available (in this context, "publicly available" means published documents which are available free of charge, or available commercially under reasonable and non-discriminatory terms to any user);*
- referenced documents which are cited only informatively as bibliographic or background material.*

A benefit of posting these files on the Unicode web site would be an easier access to these resources. Basically, except for the main text of the standard which would still be subject to the usual copyright statement enforced by ISO/IEC, the data files and the code charts would be publicly available under the usual Unicode copyright terms which allow liberal quoting.

It should also be noted that an existing large portion of the Normative References have been specified in this manner for a long time; for example, the Unicode Bidirectional Algorithm, the Unicode Normalization Forms, and many more.

As a result, the editor has created a prototype of the 5th edition of ISO/IEC 10646 using those terms. The main body of the updated standard can be accessed at <http://www.unicode.org/wg2-private/docs/n4859-5th-10646-update.pdf> (id/password required) and the other files can be accessed at <http://www.unicode.org/charts/About.html#iso10646> (or directly from the main body of the standard using the embedded links once it is posted). The link above gives terms of use allowing full usage of these charts in the context of ISO/IEC 10646:

The Unicode Consortium specifically grants ISO the right to use those 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.

Because of the substantial changes, although none change the technical content of the 5th edition, the SC2 secretariat, the Project Editor and the SC2 chair are requesting these changes to be endorsed by SC2 before publication. While these changes are substantial, they do not change in essence the technical content of the standard.

For its part, ITTF has agreed with the proposed format as long as the Unicode Consortium maintain access to these references in accordance with clause 10.2 of the Directives mentioned above.

There is some urgency as the two ongoing amendments (1 and 2) would need to be modified to take this new format into consideration.

--