

## Universal Multiple-Octet Coded Character Set International Organization for Standardization

**Doc Type:** Working Group Document

**Title:** IRG #49 Liaison Report

**Source:** Ken Lunde, Adobe Systems Incorporated

**Date:** 2017-10-20

IRG #49 took place at Adobe's HQ in San José, California, USA from **2017-10-16** through **2017-10-19**, was hosted by Adobe and Unicode, and was held in the Park conference room on the first floor of Adobe's East Tower. Adobe hosted a dinner on 2017-10-18. The main discussion topic was **IRG Working Set 2015 Version 4.0** (aka **Extension G**), in terms of dealing with comments and member body responses from the latest review cycle. The [IRG #49 Recommendations](#) (IRG N2260) are now available. Also be sure to read [IRG N2265](#) for more details.

In attendance were **30 representatives and experts** from China (9), Hong Kong SAR (1), Japan (2), ROK (4), SAT (1), TCA (3), US/Unicode (9), and Vietnam (1). Below is the group photo that was taken on the second day (2017-10-17):



Per **Recommendation IRG M49.1**, the following is the future IRG meeting schedule:

- IRG #50 Beijing, China, 2018-05-21 through 2018-05-25 (confirmed)
- IRG #51 Hanoi, Vietnam, 2018-10-22 through 2016-10-26 (confirmed)
- IRG #52 Hong Kong SAR, China, 2019-05-13 through 2019-05-17 (tentative)
- IRG #53 Chengdu/Kunming, China, 2019-10-21 through 2019-10-25 (tentative)
- IRG #54 **Seeking Host**

## **IRG Working Set 2015 (aka Extension G) Status**

Most of the meeting was spent performing IRG Working Set 2015 editorial work, which entailed going through the [Consolidated Comments](#) for Version 4.0 ([IRG N2223](#)), along with the responses from the submitters. See [IRG N2263](#) for more details.

The following **nine UTC-submitted characters** needed to be discussed in greater detail, and their final dispositions are provided below (IRG Working Set 2015 serial numbers are also provided):

- UTC-00984 00470: **Disunified**.
- UTC-00993 00829: **Disunified**.
- UTC-01075 00320: **Disunified**.
- UTC-01220 00771: **Unified with U+2D3EC** (Extension F).
- UTC-01243 02041: **Disunified**.
- UTC-01272 02262: **Disunified**.
- UTC-01276 02284: **Disunified**.
- UTC-01301 04429: **Withdrawn** per the UTC's review.
- UTC-01304 04435: **Unified with U+28B02** (Extension B) per the UTC's review.

All other UTC-submitted characters were handled according to the [UTC's response](#) to the review comments. No UTC-submitted characters require their representative glyphs to be modified.

For UK's benefit, I am listing **twenty UK-submitted characters**, with UTC source references, that needed to be discussed in greater detail, and their final dispositions are provided below:

- UTC-01349 01186: **Withdrawn** per the UK's response.
- UTC-01370 00540: **Disunified**.
- UTC-01393 02684: **Disunified**.
- UTC-01573 00198: **Unified with U+20266** (Extension B) per the UK's response.
- UTC-01639 00990: **Disunified**.
- UTC-01652 00852: **Disunified**.
- UTC-01681 00183: **Disunified**.
- UTC-01879 02752: **Disunified**.
- UTC-01950 03555: **Disunified**.

UTC-02316 04914: **Disunified**.  
UTC-02328 04871: **Disunified**.  
UTC-02339 05045: **Disunified**.  
UTC-02340 05046: **Disunified**.  
UTC-02625 00742: **Disunified**.  
UTC-02641 01113: **Disunified**.  
UTC-02770 00689: **Disunified**.  
UTC-02787 01108: **Unified with U+38C7** (Extension A) per the UK's response.  
UTC-02921 00817: **Disunified**.  
UTC-02935 02001: **Disunified**.  
UTC-02960 04752: **Disunified**.

Per **Recommendation IRG M49.9**, the prefix that is used for the source references of the UK-submitted characters will be changed from “UTC-” to “UK-,” which affects UTC-01313 through UTC-02968. This source reference change should also be reflected in [UAX #45](#), to include its data file and glyph chart, and in [UAX #38](#), in terms of the *kIRG\_USource* property.

Per **Recommendation IRG M49.2**, the following is the **IRG Working Set 2015 Version 5.0** schedule between now and IRG #50:

2017-10-27 The IRG Chief Editor distributes the Discussion Record  
2017-11-03 The IRG members' chief editors submit comments to the IRG Chief Editor  
2017-11-03 The IRG members' chief editors submit TrueType fonts to the ISO/IEC 10646 Project Editor, and updated BMP glyphs to the IRG Chief Editor  
2017-12-01 The IRG Chief Editor distributes IRG N2269 (aka IRG Working Set 2015 Version 5.0)  
2018-02-24 The IRG members' chief editors and experts submit review comments to the IRG Chief Editor  
2018-03-16 The IRG Chief Editor distributes consolidated comments  
2018-04-01 The IRG Rapporteur distributes ballot comments to IRG chief editors  
2018-05-04 The IRG members' chief editors submit responses to the consolidated comments and ballot comments to the IRG Rapporteur

Comments on metadata that effects neither unification nor ordering were skipped, and will be handled at a later date. All comments on unifications, disunifications, representative glyphs, evidence images, radical assignments, and residual strokes, which affect the repertoire and its ordering, were discussed and resolved.

## **IRG Working Set 2017 (aka Extension H) Status**

Per **Recommendation IRG M49.11**, there was an informal quality review of the preliminary version of IRG Working Set 2017 ([IRG N2227](#)) that consisted of 5,160 characters that includes submissions from China (969 characters), ROK (686 characters), SAT (323 characters), TCA (998

characters), UK (1,000 characters), UTC (193 characters), and Vietnam (991 characters). See [IRG N2264](#) for more details. As a result of the preliminary quality check, member bodies were given an opportunity to revise their submissions with the condition that characters cannot be added. The following is the schedule for the next version:

- 2018-03-16 The IRG members' chief editors submit their updated data to the IRG Rapporteur and the IRG Chief Editor
- 2018-03-30 The IRG Chief Editor distributes IRG N 2270 (aka IRG Working Set 2017 Version 0.1)
- 2018-04-20 The IRG members' chief editors and experts submit review comments
- 2018-05-04 The IRG members' chief editors submit response

Based on the discussions, I do not see a need to revise or otherwise modify the UTC's submission.

## Revised Unification Rules & IVD

Per **Recommendation WG2 M64.11** (*Review of CJK Unification Rules*) in [WG2 N4701](#), which stated “WG2 recommends that IRG reviews its CJK unification rules to minimize the number of glyph variants that are coded as separate characters,” the IRG was tasked to come up revised unification rules. Per **Recommendation IRG M49.5**, the IRG came up with two new unification rules that will apply to IRG Working Set 2017 and beyond, which also need to be reflected in [UTS #37](#) (*Unicode Ideographic Variation Database*). The following is the proposed text to be added as a new paragraph in UTS #37 Section 2, *Description*:

“In an effort to reduce the number of encoded variants, the unification rules for unified ideographs, when applied to the IVD, have been expanded to include cases whereby 1) characters that have a different structure, but whose difference is not considered significant enough to encode them as separate unified ideographs, and for which strong evidence associating them as variants of encoded characters can be provided, such as 𠂇𠂇𠂇 versus 𠂇𠂇 𠂇 (U+6E29 温) and 𠂇𠂇𠂇 versus 𠂇𠂇𠂇 (U+243B7 戴); and 2) characters with the same structure, but with different components at the second (or subsequent) level that may not be generally unifiable, and for which strong evidence associating them as variants of encoded characters can be provided, such as 𠂇𠂇𠂇 versus 𠂇𠂇𠂇 (U+8FFA 迺) and 𠂇𠂇𠂇 versus 𠂇𠂇𠂇 (U+818D 𠂇). When considering the second case, the registrant is expected to provide evidence that demonstrates 1) similarity of glyph shape; 2) use in general circulation; and 3) general acceptance as a variant.”

## Proposed Ideographic Description Characters (IDCs)

Per **Recommendation IRG M49.12**, see [L2/17-386](#) (aka IRG N2273) for a proposal to encode three new Ideographic Description Characters (IDCs), about which the IRG raised no strong objection and accepted in principle. One of the characters is binary in that two components follow it, and the other two are unary that affect single components. The latter two characters will require a change to IDS (*Ideographic Description Sequence*) syntax. The proposed code points and character names are below:



U+2FFC	IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LOWER RIGHT
U+2FFD	IDEOGRAPHIC DESCRIPTION CHARACTER HORIZONTAL AXIAL SYMMETRY
U+2FFE	IDEOGRAPHIC DESCRIPTION CHARACTER CENTRAL SYMMETRY

I personally feel that IDEOGRAPHIC DESCRIPTION CHARACTER **HORIZONTAL REFLECTION** and IDEOGRAPHIC DESCRIPTION CHARACTER **ONE HUNDRED EIGHTY DEGREE ROTATION** are better names for the second and third characters.

## Proposed Derived Simplified Character Designator (DSCD)

Dr. Lu proposed in [IRG N2274](#) the encoding of a “Derived Simplified Character Designator” (DSCD) as a way to avoid directly-encoding derived simplified ideographs. Section 6 that starts on the bottom of the first page of [IRG N2265](#) provides a detailed discussion record. Also be sure to read John Knightley’s [detailed feedback](#). Although such a mechanism could conceivably work, I generally oppose such a proposal on the grounds that it is 20 years too late: *the proverbial horse left the barn years ago*.

Also related is [IRG N1757](#) that was discussed during [IRG #36](#). The following is the IRG’s response per **Resolution IRG M36.2**: *The IRG has reviewed IRG N1757 from the UTC and is concerned that an IVD registration from the UTC might be treated as a “de facto” encoding. The IRG requests that the UTC not proceed with the registration until this issue is addressed. The IRG encourages its members to do further review of IRG N1757 and provide feedback to the UTC at any time.*

## Other Reports

Various error reports were submitted and discussed, the details of which can be found in [IRG N2265](#). More notable reports are detailed in this section.

Per [IRG N2245](#), a character that was once orphaned and to which a UAX #45 “UCI” source reference was assigned, U+24FB9 (Extension B), will longer be orphaned due to a forthcoming horizontal extension. This means that the UTC needs to decide whether to rename UCI-00942 to UTC-00942, or remove the source reference from U+24FB9 altogether. China is currently deciding about the source reference that they will use.

Per **Recommendation IRG M49.8** and [IRG N2258](#), Hong Kong SAR submitted a horizontal extension for 23 characters, and also submitted an updated font for the H-Source column of the code charts. The background is that 24 characters were added in [HKSCS-2016](#), 23 of which were ideographs. Hong Kong SAR also updated the representative glyphs to confirm to their regional standard.

Also per **Recommendation IRG M49.8** and [IRG N2268](#), Hong Kong SAR decided to move the source reference for U+2F9B2 (CJK Compatibility Ideograph), H-8FA8, to U+270F0 (Extension B). This will orphan U+2F9B2, and the UTC will need to assign a new “UCI” source reference to this character. The background is that U+2F9B2 normalizes to U+456B (Extension A), which is neither in HKSCS-2016 proper nor in its Big Five subset.

That is all.