Subject: **Preserving Variant Glyphs**
From: **Mark Davis**
Date: **2011-05-11**
Live Doc: **http://goo.gl/L2D3N**

One of the barriers to usage of NFC is the perception that it loses information. For example, here is a recent quote on an email list: " this did not seem feasible, at least not without a non-lossy normalization scheme, which Unicode currently lacks (NFC having been hijacked by the anti-compatibility-character crusades)."

If we could redo the past, we would probably have categorized the CJK compatibility characters as being compatibility equivalents and not canonical equivalents. That was recognized later on, but by then stability considerations prevented it from being changed. Excluding those characters, there are actually relatively few characters that are not allowed in NFC. However, the CJK compatibility characters are a rather broken approach to the real issue of representing variants, and a much better one has developed in the meantime, the IVD (http://www.unicode.org/ivd/). And the sequences defined by IVD *are* preserved by NFC.

To help to address the perception about NFC, we could provide IVS mappings that can be used for the most important of the compatibility CJK. This provides a mechanism for people to use to preserve the glyph distinctions that they want. By publicizing the table below, we can help address the problem. Such publication could be as a FAQ, or as something more formal with a stable location for the data table.

(Thanks to Ken Lunde and Eric Muller for supplying the data for the table below!)

```
<comp.> ;  <IVS>                 # <comp.> <ideo>
U+F91D ;  <U+6B04,U+E0100> #      欄 欄
U+F928 ;  <U+5ECA,U+E0101> #      廊 廊
U+F929 ;  <U+6717,U+E0100> #      朗 朗
U+F936 ;  <U+865C,U+E0100> #      虜 虜
U+F970 ;  <U+6BBA,U+E0100> #      殺 殺
U+F9D0 ;  <U+985E,U+E0100> #      類 類
U+F9DC ;  <U+9686,U+E0101> #      隆 隆
U+FA10 ;  <U+585A,U+E0101> #      塚 塚
```

U+FA12；<U+6674,U+E0100> #　　晴 晴

U+FA15；<U+51DE,U+E0101> #　　凞 凞

U+FA16；<U+732A,U+E0100> #　　猪 猪

U+FA17；<U+76CA,U+E0101> #　　益 益

U+FA18；<U+793C,U+E0101> #　　礼 礼

U+FA19；<U+795E,U+E0100> #　　神 神

U+FA1A；<U+7965,U+E0100> #　　祥 祥

U+FA1B；<U+798F,U+E0101> #　　福 福

U+FA1C；<U+9756,U+E0101> #　　靖 靖

U+FA1D；<U+7CBE,U+E0100> #　　精 精

U+FA1E；<U+7FBD,U+E0100> #　　羽 羽

U+FA20；<U+8612,U+E0101> #　　薒 薒

U+FA22；<U+8AF8,U+E0100> #　　諸 諸

U+FA25；<U+9038,U+E0102> #　　逸 逸

U+FA26；<U+90FD,U+E0100> #　　都 都

U+FA2A；<U+98EF,U+E0100> #　　飯 飯

U+FA2B；<U+98FC,U+E0101> #　　飼 飼

U+FA2C；<U+9928,U+E0101> #　　館 館

U+FA2D；<U+9DB4,U+E0100> #　　鶴 鶴

U+FA30；<U+4FAE,U+E0101> #　　侮 侮

U+FA31；<U+50E7,U+E0101> #　　僧 僧

U+FA32；<U+514D,U+E0101> #　　免 免

U+FA33；<U+52C9,U+E0100> #　　勉 勉

U+FA34；<U+52E4,U+E0101> #　　勤 勤

U+FA35；<U+5351,U+E0100> #　　卑 卑

U+FA36；<U+559D,U+E0101> #　　喝 喝

U+FA37 ； <U+5606,U+E0100> #　　嘆 嘆

U+FA38 ； <U+5668,U+E0101> #　　器 器

U+FA39 ； <U+5840,U+E0101> #　　塀 塀

U+FA3A ； <U+58A8,U+E0100> #　　墨 墨

U+FA3B ； <U+5C64,U+E0101> #　　層 層

U+FA3C ； <U+5C6E,U+E0100> #　　屮 屮

U+FA3D ； <U+6094,U+E0100> #　　悔 悔

U+FA3E ； <U+6168,U+E0101> #　　慨 慨

U+FA3F ； <U+618E,U+E0100> #　　憎 憎

U+FA40 ； <U+61F2,U+E0101> #　　懲 懲

U+FA41 ； <U+654F,U+E0100> #　　敏 敏

U+FA42 ； <U+65E2,U+E0100> #　　既 既

U+FA43 ； <U+6691,U+E0101> #　　暑 暑

U+FA44 ； <U+6885,U+E0100> #　　梅 梅

U+FA45 ； <U+6D77,U+E0100> #　　海 海

U+FA46 ； <U+6E1A,U+E0100> #　　渚 渚

U+FA47 ； <U+6F22,U+E0101> #　　漢 漢

U+FA48 ； <U+716E,U+E0100> #　　煮 煮

U+FA49 ； <U+722B,U+E0101> #　　爫 爫

U+FA4A ； <U+7422,U+E0100> #　　琢 琢

U+FA4B ； <U+7891,U+E0100> #　　碑 碑

U+FA4C ； <U+793E,U+E0101> #　　社 社

U+FA4D ； <U+7949,U+E0101> #　　祉 祉

U+FA4E ； <U+7948,U+E0100> #　　祈 祈

U+FA4F ； <U+7950,U+E0100> #　　祐 祐

U+FA50 ； <U+7956,U+E0101> #　　祖 祖

U+FA51； <U+795D,U+E0100> #　　祝 祝

U+FA52； <U+798D,U+E0100> #　　禍 禍

U+FA53； <U+798E,U+E0101> #　　禎 禎

U+FA54； <U+7A40,U+E0100> #　　穀 穀

U+FA55； <U+7A81,U+E0101> #　　突 突

U+FA56； <U+7BC0,U+E0101> #　　節 節

U+FA57； <U+7DF4,U+E0100> #　　練 練

U+FA58； <U+7E09,U+E0101> #　　縉 縉

U+FA59； <U+7E41,U+E0101> #　　繁 繁

U+FA5A； <U+7F72,U+E0100> #　　署 署

U+FA5B； <U+8005,U+E0101> #　　者 者

U+FA5C； <U+81ED,U+E0101> #　　臭 臭

U+FA5D； <U+8279,U+E0101> #　　艹 艹

U+FA5E； <U+8279,U+E0102> #　　艹 艹

U+FA5F； <U+8457,U+E0101> #　　著 著

U+FA60； <U+8910,U+E0101> #　　褐 褐

U+FA61； <U+8996,U+E0101> #　　視 視

U+FA62； <U+8B01,U+E0100> #　　謁 謁

U+FA63； <U+8B39,U+E0100> #　　謹 謹

U+FA64； <U+8CD3,U+E0100> #　　賓 賓

U+FA65； <U+8D08,U+E0101> #　　贈 贈

U+FA66； <U+8FB6,U+E0101> #　　辶 辶

U+FA67； <U+9038,U+E0101> #　　逸 逸

U+FA68； <U+96E3,U+E0100> #　　難 難

U+FA69； <U+97FF,U+E0101> #　　響 響

U+FA6A； <U+983B,U+E0100> #　　頻 頻

U+FA6B ; <U+6075,U+E0101> #　　惠 惠

U+FA6C ; <U+242EE,U+E0101> #　熙 熙

U+FA6D ; <U+8218,U+E0101> #　　袨 舘