

## Remove ambiguity from D14 Noncharacters for Unicode 16.0

Date: August 31

Submitted by: asmus@unicode.org

**Problem: the word “internal” is ambiguous in definition D14 *Noncharacters*.**

In reviewing an IETF internet draft I came across language that cited the “reserved for internal use” language from D14. However, without the context of text passages in §2.13 “Special Character” or §23.7 “Noncharacters” the meaning of this reservation is not unambiguous to the reader of that internet draft.

In the context of a body of work defining protocols (like the IETF RFCs) “reserved for internal use” may refer to items that are reserved to be used solely for protocol internal purposes, whereas the sense in Unicode encompasses use that is internal to a system or application (or protocol) that implements the Unicode Standard. Instead of noncharacters being somehow for internal use in the Unicode Standard, they are internal in the sense of not being intended for interchange.

The existing language does not make that distinction, or, in other words, does not prevent a reader from applying their understanding of the word “internal”, leading to an ambiguity.

*D14 Noncharacter: A code point that is permanently reserved for internal use. Noncharacters consist of the values U+nFFFE and U+nFFFF (where n is from 0 to 10 16) and the values U+FDD0..U+FDEF.*

with one of the bullets just repeating the statement without any help in disambiguation

- These code points are permanently reserved as noncharacters

Other text in the Standard spells out the key features of noncharacters in different ways that make the intent much clearer and rule out alternative interpretation of the word “internal”. Those statements describe noncharacters as follows:

1. They are code points that can never be assigned to abstract characters
2. They are not intended for interchange
3. Applications may (freely) use them for application internal purposes
4. Some may be suitable as sentinel values
5. If encountered outside of an implementation they have no interpretable semantics other than their status as noncharacter code point.

### Preferred Solution

The preferred fix would be to update the language of D14 as follows:

*D14 Noncharacter: A code point that is permanently reserved and will never be assigned to an abstract character. Noncharacters consist of the values U+nFFFE and U+nFFFF (where n is from 0*

*to 10 16) and the values U+FDD0..U+FDEF. They are not intended for interchange, but may be used by an implementation for internal purposes.*

- Possible use cases include application internal sentinel values
- For more information, see Section 23.7, *Noncharacters*.

**Alternatives:**

- A. Same as preferred solution, but with the final sentence moved into a bullet
- B. Same as existing, but with new bullets covering items (1) through (5) added and the existing duplicative bullet removed.

**Glossary**

Change the glossary item to match the suggested language for the “preferred solution” instead of repeating the current definition with its ambiguous use of the word “internal”.