

# Stabilize Technical Reports #6, #16, #26

Markus Scherer, 2017-nov-24

I propose that we stabilize

- [UTS #6 A Standard Compression Scheme for Unicode](#)
- [UTR #16 UTF-EBCDIC](#)
- [UTR #26 Compatibility Encoding Scheme for UTF-16: 8-Bit \(CESU-8\)](#)

like we stabilized [UTS #22 Unicode Character Mapping Markup Language](#) before.

This is to indicate that these reports are not in active development, without marking them as superseded or withdrawn. It should also serve to de-emphasize them for potential new users, and reduce the number of active specifications to learn about for new visitors to the [Technical Reports](#) page.

**SCSU** defines a compact encoding, which is sometimes useful. However, Unicode text is much more commonly stored and transmitted in UTF-8 which is less compact (except for ASCII), much simpler, and does not present any security issues. For longer texts, general-purpose compression is effective and common.

**UTF-EBCDIC** is an encoding form similar to UTF-8, but based on EBCDIC instead of ASCII. From [Wikipedia](#): “This encoding form is rarely used, even on the EBCDIC-based mainframes for which it was designed. IBM EBCDIC-based mainframe operating systems, such as z/OS, usually use UTF-16 for complete Unicode support. For example, DB2 UDB, COBOL, PL/I, Java and the IBM XML toolkit support UTF-16 on IBM mainframes.”

From UTR #26: “**CESU-8** defines an encoding scheme for Unicode identical to UTF-8 except for its representation of supplementary characters. In CESU-8, supplementary characters are represented as six-byte sequences resulting from the transformation of each UTF-16 surrogate code unit [...] without first converting the input surrogate pairs to a scalar value.” The report itself says that it “is intended for internal use” and “It is not intended nor recommended as an encoding used for open information exchange. The Unicode Consortium, does not encourage the use of CESU-8, but does recognize the existence of data in this encoding ...”