

Unicode liaison report to ISO/IEC JTC 1/SC 22 #42

To: ISO/IEC JTC 1/SC 22
From: Robin Leroy (eggrobin@unicode.org)
Date: 2023-08-01

The liaison representative has worked with the Ada Rapporteur Group (under WG 9) and SG 16 (under WG 21) to assist in interpreting and applying Unicode standards, and welcomes questions from other Working Groups, should such questions arise in their work.

In the other direction, several dispositions taken by the Unicode Technical Committee were informed by feedback from the liaison regarding SC 22 projects: see [176-A59](#), [176-A63](#), and [176-C18](#), and the documents referenced therein.

Further, the upcoming Unicode version has changes relevant to SC 22.

Unicode Version 15.1 and Unicode Technical Standard #55 Version 1

The Unicode Consortium will publish Unicode Version 15.1 in September 2023. In this version, Unicode Standard Annex #31, “Unicode Identifiers and Syntax”, undergoes a major revision.

The content of that update is informed by difficulties encountered by programming language designers inside and outside SC 22. It includes improved guidance on the interpretation of whitespace, including line terminators and ignorable format control characters; as well as new advice on compatibility considerations, as changing requirements may require a programming language to switch from one identifier definition to another.

Simultaneously with the 15.1 release, a new standard, Unicode Technical Standard #55, “Unicode Source Code Handling”, will be published. Unicode Technical Standard #55 provides recommendations for programming environments; it also incorporates a high-level guide to the normative definitions given in Unicode Standard Annex #31, clarifying which options are appropriate depending on the programming language.

One of the major goals during the development of these standards was to provide guidance on addressing spoofing and usability concerns arising from the usage of bidirectional script and scripts with lookalike characters. The recommended mitigations are mainly applicable to source code editors, tooling such as linters and pretty-printers, or compiler warnings, which are typically outside the scope of language specifications. However, language specifications should follow certain recommendations so that the other tools can apply their mitigations effectively, hence the changes to UAX #31 and the guidance for computer language specifications in UTS #55. The recommendations applicable to language specifications are relevant even when spoofing is not a concern; they are general best practices for identifier definitions, whitespace characters, and syntax characters.

Additional changes were informed by issues encountered by the languages under the aegis of SC 22; notably, the addition of KATAKANA MIDDLE DOT to XID_Continue resolves a mistake made in Unicode Version 4.1 (2005), which was noted and worked around in the COBOL 2023 standard (ISO/IEC 1989:2023).

Until publication of these Unicode standards in September, 2023, see [Proposed Update UAX #31](#), [Draft UTS #55](#).