

# Modern Network Unicode

[draft-bormann-dispatch-modern-network-unicode-04](#)

Carsten Bormann, ALLDISPATCH 2024-03-18

# History

- October 1969: RFC20, "[ASCII](#) format for Network Interchange", V. Cerf. (Also [STD80](#))
- May 1983: RFC854, "TELNET PROTOCOL SPECIFICATION", J. Postel, J. Reynolds. (defined "**Network Virtual Terminal**", NVT)
- January 1998: RFC2277, "[IETF Policy on Character Sets and Languages](#)". H. Alvestrand. (Also BCP18)
- November 2003: RFC3629, "[UTF-8](#), a transformation format of ISO 10646". F. Yergeau. (Also STD63)
- March 2008: RFC5198, "Unicode Format for Network Interchange", J. Klensin, M. Padlipsky. (defined "Unicode NVT")

# The Vacuum

- Unicode didn't need an NVT that much any more.  
A Unicode teletype print head would be huge, anyway.
  - Protocols now have structure (XML, JSON, CBOR, ...)  
Application protocols define pieces of text (**text strings**) within that structure
  - RFC 5198 was still useful, but not a good fit
  - Unicode is more **complex** than ASCII was!
- **Protocol designers need help**

# Writing up Modern Network Unicode

Triggered by work on COSWID

Started July 2019 on art@, with excursions to i18ndir@

Lots of good feedback,

including feedback solicited from W3C I18N contributors

- Simple menu of choices, each with (inevitable) options
- Appendices (0.5 to 1.5 pages each) on
  - Terminology
  - History/Legacy issues
  - Unicode Normalization
  - Relationship to RFC 5198

# Objective: Be useful

- as a **guide** for protocol developers
- as a **reference** a new specification can point to