

Network Working Group  
Internet-Draft  
Updates: [5234](#) (if approved)  
Intended Status: Standards Track  
Expires: April 20, 2016

S. Leonard  
Penango, Inc.  
October 18, 2015

**Comprehensive Core Rules and References for ABNF**  
**draft-seantek-abnf-more-core-rules-04**

Abstract

This document extends the base definition of ABNF (Augmented Backus-Naur Form) to include comprehensive support for certain symbols related to ASCII, and defines a reference syntax.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 20, 2016.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## 1. Introduction

Augmented Backus-Naur Form (ABNF) [[RFC5234](#)] is a formal syntax that is popular among many Internet specifications. Many Internet documents employ this syntax along with the Core Rules defined in [Appendix B.1 of \[RFC5234\]](#). However, the Core Rules do not specify many symbols in the ASCII range that are also needed by these relying documents, forcing document authors to define them as local rules. Sometimes different documents define these common symbols in different ways, resulting in confusion or incompatibility when the rules are misread or are combined with other sets of rules. Furthermore, [[RFC5234](#)] does not clarify whether referencing [[RFC5234](#)] for ABNF automatically defines its Core Rules.

[RFC5234] also lacks a syntax for referring to rules from other specifications. Instead, authors have been required to name the rules and sources in the specification prose. While this method has served authors well, it has hampered machine-readable ABNF efforts for services such as syntax highlighting, automatic grammar checking, and compiling into target computer languages.

## 2. Comprehensive Core Rule Update

This document provides Core Rules that include comprehensive support for certain symbols, namely DELETE (DEL) and the C0 controls in [[ASCII86](#)], which for purposes of this document is equivalent to [[RFC0020](#)].

## 3. Reference Syntax

To reference a rule in another ABNF grammar, use the syntax `rulename@REF`. The referenced rule resolves to terminal values in the context of the referenced ABNF grammar, essentially replacing the verbiage: "{RULE} is taken from [{Section X} of] {[RFCXXXX]}" in prose specification text that introduces the ABNF. The following enhancement to [[RFC5234](#)] permits this referenced-rule syntax as an incremental element:

```
element      =/ refrule

refrule      =  rulename "@" ruleref

ruleref      =  ref-doc / ref-uri

ref-doc      =  "[" 1*(%x20-5A / %x5C / %x5E-7E) "]"
```



```

; NB: the part after ":" could be 1*VCHAR for even greater simplicity
ref-uri      = ALPHA *( ALPHA / DIGIT / "+" / "-" / "." )
               ":" 1*( "!" / %x23-3B / "=" / %x3F-5B /
                       "]" / "_" / %x61-7A / "~" )

```

In the referenced-rule production (refrule), the rulename production preceding the "@" specifies the name of the rule in the reference containing ABNF. The ruleref production following the "@" specifies the reference containing ABNF. The precise manner in which the reference is associated with the grammar contained therein is not defined by this specification. Furthermore, this specification does not define the semantics if a rule is found in a grammar that is not ABNF. (This limitation is because rule names in ABNF are case-insensitive and drawn from a limited character repertoire. Some rule names in other BNFs may be unreachable or ambiguous, even though the productions named by the rules are linguistically compatible.)

The form ref-doc is a document reference of a resource containing ABNF. The term "document reference" refers to "the document containing this ABNF (i.e., the instance of these production rules)". In IETF-related publications, ref-doc conveniently is of the same form as document references, such as "[[RFC1605](#)]".

The form ref-uri is supposed to be a Uniform Resource Identifier [[RFC3986](#)], but this specification imposes no requirement to validate conformance to the URI production of [[RFC3986](#)]. The reference is to the resource containing ABNF identified by the URI. Fragment components might be present, but only if the resource defines the fragment to mean a range of text (i.e., not just a point in the text).

Stylistically, authors are encouraged to put reference syntax at the top of a list of rules, and to limit usage of the reference syntax to the single element of a rule definition. For example:

```

You      = Edward@[FFIV]
spoony   = spoony@[FFIV]
bard     = bard@[FF-JOB-CLASS]
chara    = Tellah@[FFIV]

insult   = chara ":" You spoony bard "!"

```

#### 4. Effects on [RFC 5234](#)

Formally, this document updates [[RFC5234](#)] but does not modify it in situ. Authors need to reference this document if they want to include these enhancements; bare references to [[RFC5234](#)] do not include this specification (or, for that matter, [[RFC7405](#)]). This directive



follows a model whereby document authors can choose whether to invoke particular enhancements to ABNF. As time goes on, the IETF can determine how often these enhancements are invoked, and can decide whether to include them as part of a revision to the base [[RFC5234](#)].

A bare reference to this document invokes the reference syntax enhancement as well as the Core Rules of [Appendix A](#) (i.e., the Core Rules do not have to use reference syntax). Nevertheless, document authors are free to qualify a reference to this document to invoke each feature selectively.

[Appendix A](#) of this document is meant to mirror [Appendix B.1 of \[RFC5234\]](#); therefore, concurrently referencing [Appendix B.1 of \[RFC5234\]](#) is superfluous. Document authors who reference this document should use the rules of [Appendix A](#), and should not attempt to redefine or provide incremental alternatives to them (except for backwards compatibility with prior documents).

## **5. IANA Considerations**

This document implies no IANA considerations.

## **6. Security Considerations**

Security is truly believed to be irrelevant to the Core Rules in this document.

Unfortunately security is relevant to the reference syntax in this document. Using the reference syntax facilitates automated processing of ABNF. A malicious source could supply different ABNF as an attack vector on a compiled program. Furthermore, referring to a mutable resource (e.g., a document series such as BCP) permits the resource to change its contained ABNF, which may be well-intentioned but have side-effects when combined with the referring ABNF. Authors should stick to persistent, durable references, whose integrity can be validated easily.

## **7. References**

### **[7.1. Normative References](#)**

- [ASCII86] American National Standards Institute, "Coded Character Set -- 7-bit American Standard Code for Information Interchange", ANSI X3.4, 1986.
- [RFC0020] Cerf, V., "ASCII format for network interchange", [RFC 20](#), October 1969.



[RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, [RFC 5234](#), January 2008.

## **[7.2. Informative References](#)**

[UNICODE] The Unicode Consortium, "The Unicode Standard, Version 8.0.0", The Unicode Consortium, August 2015.

[RFC1345] Simonsen, K., "Character Mnemonics and Character Sets", [RFC 1345](#), June 1992.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, [RFC 3986](#), January 2005.

[RFC5198] Klensin, J. and M. Padlipsky, "Unicode Format for Network Interchange", [RFC 5198](#), March 2008.

## **[Appendix A. Comprehensive Core Rules](#)**

Certain basic rules are in uppercase, such as SP, HTAB, CRLF, DIGIT, ALPHA, etc.

ALPHA	=	%x41-5A / %x61-7A	; A-Z / a-z
BIT	=	"0" / "1"	
CHAR	=	%x01-7F	; any 7-bit US-ASCII character, ; excluding NUL
CR	=	%x0D	; carriage return
CRLF	=	CR LF	; Internet standard newline
CTL	=	%x00-1F / %x7F	; controls
DIGIT	=	%x30-39	; 0-9
DQUOTE	=	%x22	; " (Double Quote)
HEXDIG	=	DIGIT / "A" / "B" / "C" / "D" / "E" / "F"	





HTAB	= %x09	; horizontal tab
LF	= %x0A	; linefeed
LWSP	= *(WSP / CRLF WSP)	; Use of this linear-white-space rule ; permits lines containing only white ; space that are no longer legal in ; mail headers and have caused ; interoperability problems in other ; contexts. ; Do not use when defining mail ; headers and use with caution in ; other contexts.
OCTET	= %x00-FF	; 8 bits of data
SP	= %x20	
VCHAR	= %x21-7E	; visible (printing) characters
WSP	= SP / HTAB	; white space
NUL	= %d0	
SOH	= %d1	
STX	= %d2	
ETX	= %d3	
EOT	= %d4	
ENQ	= %d5	
ACK	= %d6	
BEL	= %d7	
BS	= %d8	
HT	= %d9	; also defined as HTAB
VT	= %d11	
FF	= %d12	; (literally used in every RFC)
S0	= %d14	
SI	= %d15	
DLE	= %d16	
DC1	= %d17	
DC2	= %d18	
DC3	= %d19	



DC4	=	%d20
NAK	=	%d21
SYN	=	%d22
ETB	=	%d23
CAN	=	%d24
EM	=	%d25
SUB	=	%d26
ESC	=	%d27
FS	=	%d28
GS	=	%d29
RS	=	%d30
US	=	%d31
DEL	=	%d127
ASCII	=	%x00=7F
C0	=	%x00-1F

## [Appendix B](#). Guidance for Rule Names for C1 Controls and Other Desiderata

Internet protocols have been migrating to Unicode and specifically UTF-8 for general text encoding. Authors need to consider the presence and possible effects of characters and code points beyond ASCII. See [\[RFC5198\]](#). Therefore, the following rule names may take on special meanings. This document does not formally define these rule names, nor does this document prohibit other specifications from using them. However, authors ought only to use these rule names in their normal and natural senses. For the underlying sources, consult [\[UNICODE\]](#) and [\[RFC1345\]](#).

ABNF rules resolve into a string of terminal values. Such a value "is merely a non-negative integer"; only context can furnish a specific mapping of values into a character set. Therefore, even if Unicode is specified, mappings between terminal values beyond %x7F may result in different bit combinations depending on the encoding method.

This document does not purport to change the character set of ABNF itself, which remains [\[ASCII86\]](#). (See [\[RFC5234\]](#).)

UNICODE	<U+0000-U+D7FF / U+E000-U+10FFFF>
BEYONDASCII	<U+0080-U+D7FF / U+E000-U+10FFFF>
	[[DISCUSS: these definitions are limited to
	the Unicode scalar values.]]
C1	<U+0080-U+009F>
PAD	<U+0080>
HOP	<U+0081>
BPH	<U+0082>



NBH	<U+0083>
IND	<U+0084>
NEL	<U+0085>
NL	<possibly CRLF, CR, LF, NEL, or any combination thereof, but not LS or PS>
SSA	<U+0086>
ESA	<U+0087>
HTS	<U+0088>
HTJ	<U+0089>
VTs	<U+008A>
PLD	<U+008B>
PLU	<U+008C>
RI	<U+008D>
SS2	<U+008E>
SS3	<U+008F>
DCS	<U+0090>
PU1	<U+0091>
PU2	<U+0092>
STS	<U+0093>
CCH	<U+0094>
MW	<U+0095>
SPA	<U+0096>
EPA	<U+0097>
SOS	<U+0098>
SGCI	<U+0099>
SCI	<U+009A>
CSI	<U+009B>
ST	<U+009C>
OSC	<U+009D>
PM	<U+009E>
APC	<U+009F>
NBSP	<U+00A0>
SHY	<U+00AD>
LS	<U+2028>
PS	<U+2029>

## Author's Address

Sean Leonard  
Penango, Inc.  
5900 Wilshire Boulevard  
21st Floor  
Los Angeles, CA 90036  
USA

EMail: [dev+ietf@seantek.com](mailto:dev+ietf@seantek.com)

URI: <http://www.penango.com/>

