



XMPP

XEP-0029: Definition of Jabber Identifiers (JIDs)

Craig Kaes

<mailto:craigk@jabber.com>

<xmpp:craigk@jabber.com>

2022-09-23

Version 1.1.1

Status	Type	Short Name
Retracted	Standards Track	N/A

Note well: this document was superseded by RFC 3920 and RFC 6122, which in turn have been superseded by RFC 7622. This document defines the exact nature of a Jabber Identifier (JID).

Legal

Copyright

This XMPP Extension Protocol is copyright © 1999 – 2024 by the [XMPP Standards Foundation](#) (XSF).

Permissions

Permission is hereby granted, free of charge, to any person obtaining a copy of this specification (the "Specification"), to make use of the Specification without restriction, including without limitation the rights to implement the Specification in a software program, deploy the Specification in a network service, and copy, modify, merge, publish, translate, distribute, sublicense, or sell copies of the Specification, and to permit persons to whom the Specification is furnished to do so, subject to the condition that the foregoing copyright notice and this permission notice shall be included in all copies or substantial portions of the Specification. Unless separate permission is granted, modified works that are redistributed shall not contain misleading information regarding the authors, title, number, or publisher of the Specification, and shall not claim endorsement of the modified works by the authors, any organization or project to which the authors belong, or the XMPP Standards Foundation.

Warranty

NOTE WELL: This Specification is provided on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE.

Liability

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall the XMPP Standards Foundation or any author of this Specification be liable for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising from, out of, or in connection with the Specification or the implementation, deployment, or other use of the Specification (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if the XMPP Standards Foundation or such author has been advised of the possibility of such damages.

Conformance

This XMPP Extension Protocol has been contributed in full conformance with the XSF's Intellectual Property Rights Policy (a copy of which can be found at <https://xmpp.org/about/xsf/ipr-policy>) or obtained by writing to XMPP Standards Foundation, P.O. Box 787, Parker, CO 80134 USA).

Contents

1	Introduction	1
2	JIDs	1
2.1	Grammar	1
2.2	Domain Identifier	2
2.3	Node Identifier	2
2.4	Resource Identifier	2
2.5	Limited Resources	3

1 Introduction

Note: This document was superseded by [RFC 3920](#)¹ and [RFC 6122](#)², which in turn have been superseded by [RFC 7622](#)³.

Jabber Identifiers (JIDs) uniquely identify individual entities in the Jabber network. To date, their syntax has been defined by convention, existing implementations, and available documentation. As it exists, certain characters that are allowed in JIDs cause ambiguity, and the lack of a size limit on resources defies database schemas and causes some trivial JID operations to require dynamic memory allocation. This document seeks to both define and improve the existing JID syntax. This document will not explain the general usage or nature of JIDs, instead focusing on syntax.

2 JIDs

JIDs consist of three main parts:

1. The node identifier (optional)
2. The domain identifier (required)
3. The resource identifier (optional)

JIDs are encoded UTF-8. A grammar will be presented first, followed by specific clarifying and further restricting remarks.

2.1 Grammar

```

<JID> ::= [<node>"@"]<domain>["/"<resource>]
<node> ::= <conforming-char>[<conforming-char>]*
<domain> ::= <hname>["."<hname>]*
<resource> ::= <any-char>[<any-char>]*
<hname> ::= <let>|<dig>[[<let>|<dig>|"-"]*<let>|<dig>]
<let> ::= [a-z] | [A-Z]
<dig> ::= [0-9]
<conforming-char> ::= #x21 | [#x23-#x25] | [#x28-#x2E] |
                    [#x30-#x39] | #x3B | #x3D | #x3F |
                    [#x41-#x7E] | [#x80-#xD7FF] |
                    [#xE000-#xFFFF] | [#x10000-#x10FFFF]

```

¹RFC 3920: Extensible Messaging and Presence Protocol (XMPP): Core <<http://tools.ietf.org/html/rfc3920>>.

²RFC 6122: Extensible Messaging and Presence Protocol (XMPP): Address Format <<http://tools.ietf.org/html/rfc6122>>.

³RFC 7622: Extensible Messaging and Presence Protocol (XMPP): Address Format <<http://tools.ietf.org/html/rfc7622>>.

```
<any-char> ::= [#x20-#xD7FF] | [#xE000-#xFFFD] |  
                [#x10000-#x10FFFF]
```

2.2 Domain Identifier

A domain identifier is a standard DNS hostname as specified in RFC952 ⁴ and RFC1123. ⁵ It is case-insensitive 7-bit ASCII and limited to 255 bytes. It is the only required component of a JID.

2.3 Node Identifier

Node identifiers are restricted to 256 bytes, They may contain any Unicode character higher than #x20 with the exception of the following:

1. #x22 (")
2. #x26 (&)
3. #x27 (')
4. #x2F (/)
5. #x3A (:)
6. #x3C (<)
7. #x3E (>)
8. #x40 (@)
9. #x7F (del)
10. #xFFFE (BOM)
11. #xFFFF (BOM)

Case is preserved, but comparisons will be made in case-normalized canonical form.

2.4 Resource Identifier

Resources identifiers are case-sensitive and are limited to 256 bytes. They may include any Unicode character greater than #x20, except #xFFFE and #xFFFF.

⁴<http://www.ietf.org/rfc/rfc952.txt>

⁵<http://www.ietf.org/rfc/rfc1123.txt>

2.5 Limited Resources

To date, resource identifiers have not had a fixed limit on their length. This document seeks to limit it to 256 bytes for the following reasons:

1. In order to perform JID manipulations safely, one cannot use stack space if there is no limit. This forces temporary calculations onto the heap which is unnecessarily costly.
2. As a fixed length character field, a resource identifier is more easily stored in, searched on, and retrieved from a database. If an end user may store an encyclopedia's worth of information in their resource, then the only way it can be stored without truncating it is to store it as a large object (BLOB or CLOB). Depending on the database used, that makes it either grossly inefficient or impossible to search on.
3. There exist denial of service attacks stemming from an unlimited resource length.

In a worst-case encoding, such as Han ideographs, 256 bytes will provide enough storage space for 64 character points. This provides a lower bound on the number of characters a node may have in its resource.

Specifying limits in terms of bytes instead of characters is somewhat arbitrary once a lower bound for characters is established. This document proposes limits in terms of bytes mainly because doing so results in parsing efficiency; specifically, an implementation does not have to un-encode the UTF-8 string for the sole purpose of further restricting character sets that require fewer than four bytes per character point. It is sufficient to have a lower bound on characters and an upper bound on bytes.