

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 02, 2013

J. Levine
Taughannock Networks
P. Hoffman
VPN Consortium
October 2012

Variants in Second-Level Names Registered in Top Level Domains
draft-levine-tld-variant-01

Abstract

IDNA [[RFC5890](#)] provides a method to map a subset of names written in Unicode into the DNS. Some languages allow a particular name to be written in multiple ways that are represented differently in IDNA, known as "variants". This document surveys the approaches that ICANN-contracted top level domains have taken to the registration and provisioning of variant names. This document is not (and will not be) a product of the IETF, and does not (and will not) propose any method to make variants work "correctly".

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 02, 2013.

Copyright Notice

Copyright (c) 2012 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.

Table of Contents

1.	Introduction	2
2.	Terminology	3
3.	Base documents	4
4.	Domain practices	4
4.1.	AERO	4
4.2.	ASIA	4
4.3.	BIZ	4
4.4.	CAT	5
4.5.	COM	5
4.6.	COOP	5
4.7.	INFO	5
4.8.	JOBS	6
4.9.	MOBI	6
4.10.	MUSEUM	6
4.11.	NAME	6
4.12.	NET	6
4.13.	ORG	6
4.14.	POST	6
4.15.	PRO	6
4.16.	TEL	6
4.17.	TRAVEL	7
4.18.	XXX	7
5.	Note about the references REMOVE BEFORE PUBLICATION	7
6.	References	7
	Authors' Addresses	9

[1.](#) Introduction

IDNA [[RFC5890](#)] provides a method to map a subset of names written in Unicode into the DNS [[RFC1035](#)]. Some languages allow a particular name to be written in multiple ways that are represented differently in IDNA, known as "variants". In some cases, the variants are multiple equally valid ways of writing the same thing, such as traditional and simplified Chinese characters. Some languages written in Latin characters with accents and diacritical marks, known as decorated characters, allow the decorations to be omitted in some situations, such as French which often omits accents on capital letters. Due to the difficulty of representing decorated characters in ASCII systems, many users have informally used undecorated characters in DNS names, even when they are not linguistically equivalent to the decorated versions.

The proper handling of variant names has been a topic of extensive debate and research, with little consensus reached on how to handle them, or even what characters are variants of each other. Many people would like variant names to behave "the same", for a diverse range of meanings of "same." In some cases it is a textual similarity, such as variants having corresponding DNS records, in some it is functional similarity, such as variant names resolving to the same web server, or the same page in a web server, while in others it is user experience similarity, such as names resolving to web pages which while not identical are perceived by human users as equivalent.

This document provides a snapshot of variant handling in the top level domains managed by ICANN, so called gTLDs (generic TLDs) and sTLDs (sponsored TLDs), as of late 2012. We chose those domains because ICANN requires each TLD to describe its IDN and variant practices, and the TLD zone files are available for inspection, to verify what actually goes into the zones.

The authors note that ICANN has no agreed-on definition of "variant". Since "variant" can mean vastly different things to different people, there is also no agreement about when when two zones are supposed to "behave the same". Also, the gTLDs and sTLDs might have different views of what variants are and are not required to report to ICANN about their policies.

2. Terminology

We use some terminology that has become fairly well agreed when discussing variant names.

Bundle: The IDN practices documents (see below) can identify sets of characters that are considered equivalent using Language Variant Tables, defined in [[RFC3743](#)]. A set of names in which the characters in each position are equivalent is known as bundle, or more technically as an IDL Package. The variant rules vary among languages, and for the same language can vary among TLDs. Many languages, including most written in Latin script, do not define equivalent characters, and hence do not have bundles.

Preferred variant: When a Language Variant Table defines sets of equivalent characters, one character in each set is designated as preferred. In a bundle, the variant that consists entirely of preferred characters is the preferred variant. Typically it is the variant that best matches the way that words are written in natural language. Preferred variants are both language and country specific. For example, in some Chinese-speaking countries, the preferred variant is simplified characters, while

in others it is traditional characters.

Blocking: When one name in a bundle is registered in a TLD, the rest of the names in the bundle are often blocked, meaning that nobody can register them. In some cases even though the names are blocked from registration by anyone else, the registrant or registry can activate some or all of the otherwise blocked names.

Parallel NS: Multiple names in a bundle are provisioned in the TLD with identical NS records, so they all are handled by the same name servers.

DNAME aliasing: The DNAME [[RFC6672](#)] DNS record creates a shadow tree of DNS records, roughly as though there were a CNAME in the shadow tree pointing to each name in the target tree. DNAMEs have been used both as second-level names, to provide resolution for several names in a bundle, and as first-level names, to provide resolution for every name under a TLD.

[3.](#) Base documents

ICANN has published a variety of documents on variant management. The most important are the "Guidelines for the Implementation of Internationalized Domain Names" issued in Version 1.0 [[G1](#)] and Version 3.0 [[G3](#)].

TLDs are supposed to register an IDN practices document with IANA for each language in which the TLD accepts IDN registrations, to be entered in an IANA registry [[IANAIDN](#)]. The practices document lists the Unicode characters allowed in names in the language, which characters are considered equivalent, and which of an equivalent group is preferred. Some TLDs have been more diligent than others at keeping the registry up to date.

Some of the ICANN agreements with each TLD [[ICANNAGREE](#)] describe the TLD's IDN practices, but most don't.

[4.](#) Domain practices

[4.1.](#) AERO

The .AERO TLD has no IDNs, and no rules or practices for them.

[4.2.](#) ASIA

The .ASIA domain accepts registrations in many Asian languages. They have IANA tables for Japanese, Korean, and Chinese. The IANA tables refer to their CJK IDN policies [[ASIACJK](#)], which say that applied-for and preferred IDN variants are "active and included in the zone." No IDN publication mechanism is described in the documentation, but since the zone file contains no DNAMEs, they must be using parallel

NS for variants.

4.3. BIZ

Levine & Hoffman

Expires April 02, 2013

[Page 4]

ICANN gave the registry (Neustar) non-specific permission to register in a letter in 2004 [[TWOMEY04A](#)]. The IDN rules were apparently discussed with ICANN, but not defined; see Appendix 9 of the registry agreement [[ICANNBIZ9](#)].

They have about a dozen IANA tables. No IDN publication mechanism is described, but from inspection it appears that variants are blocked.

[4.4.](#) **CAT**

The IDN rules are described in [Appendix S](#) Part VII.2 [[ICANNCATS](#)] of the ICANN agreement. "Registry will take a very cautious approach in its IDN offerings. IDNs will be bundled with the equivalent ASCII domains." The only language is Catalan. No IDN publication mechanism is described.

Although the Catalan IDN practices document does not identify variant characters, in practice bundles consist of names with accented and unaccented vowels, and "ll" and the Catalan "ela geminada" written as two L's with a dot in between.

When a registrant registers an IDN, the registry also includes the ASCII version. From inspection of the zonefile, the ASCII version is provisioned with NS, and the IDN is a DNAME alias of the ASCII version.

[4.5.](#) **COM**

ICANN and Verisign have extensive correspondence about IDNs and variants, including letters to ICANN from Ben Turner [[TURNER03](#)] and Ed Lewis [[LEWIS03](#)].

The IANA registry has tables for several dozen languages, including archaic languages such as hieroglyphics and Aramaic. Verisign publishes documents describing Scripts and Languages [[VRSNLANG](#)], Character Variants [[VRSNCHAR](#)], Registration Rules [[VRSNRULES](#)], and additional registration logic [[VRSNADDL](#)].

In Chinese, variants are blocked (see [[VRSNADDL](#)].) In other languages there appears to be no bundling or blocking.

[4.6.](#) **COOP**

The .COOP TLD has no IDNs, and no rules or practices for them.

[4.7.](#) **INFO**

The IANA registry has tables for Danish, Hungarian, Lithuanian, Latvian, and Swedish from 2005. The domain also has names in Greek,

Russian, Arabic, and other languages but no IANA tables.

The registry agreement Appendix 9 [[ICANNINF09](#)] refers to a 2003

letter from Paul Twomey [[TWOMEY03](#)] that refers to blocking variants.

[4.8.](#) **JOBS**

The .JOBS TLD has no IDNs, and no rules or practices for them.

[4.9.](#) **MOBI**

The zone file has about 22,000 IDNs. The domain has no tables at IANA. The registry agreement [Appendix S](#) [[ICANNMOBIS](#)] says that IDNs are provisioned according to [[G1](#)].

[4.10.](#) **MUSEUM**

The zone file has many IDNs, but spot checks find that many are lame or dead. A 2004 letter from Paul Twomey [[TWOMEY04](#)] refers to [[G1](#)].

The registry has a detailed policy page [[MUSEUMIDN](#)]. IDNs are accepted in Latin and Hebrew scripts, with plans for Arabic, Chinese, Japanese, Korean, Cyrillic, and Greek. They do no bundling or blocking, but names that may be confusable due to visual similarity are not allowed, apparently determined by manual inspection, which is practical due to the very small size of the domain.

[4.11.](#) **NAME**

The .NAME domain is now managed by Verisign, and has same long list of scripts as .COM and .NET. A 2004 letter from Paul Twomey [[TWOMEY04B](#)] refers to [Appendix K](#) of the agreement, but appendices are numbered. Appendix 11 [[ICANNNAME11](#)] is about restrictions on names, but says nothing about IDNs. The Letter above refers to [[G1](#)].

[4.12.](#) **NET**

The domain is managed the same as .COM.

[4.13.](#) **ORG**

A 2003 letter from Paul Twomey [[TWOMEY03A](#)] refers to [[G1](#)]. The registry has a list of IDN languages [[PIRIDN](#)], all written in Latin script. The practices for some but not all are registered with IANA, Since none of the languages do bundling, there is presumably no blocking.

[4.14.](#) **POST**

The .POST TLD appears to have no registrations at all yet.

[4.15.](#) **PRO**

The .PRO TLD has no IDNs, and no rules or practices for them.

[4.16.](#) TEL

The zone has many IDNs. It is probably operating according to a 2004 letter from Paul Twomey [[TWOMEY04A](#)] to Neustar which did not mention specific TLDs. Its policy page [[TELPOLICY](#)] has links to IDN practices for 17 languages, all but one of which are registered with IANA. None of the Latin scripts do bundling or blocking. The Japanese practices say that variants are blocked. The Chinese practices document says:

Therefore, in addition to the blocking mechanism, bundling is also implemented for the Chinese language IDNs. When registering a Chinese language IDN (primary domain name) up to two additional variant domain names will be automatically registered. The first variant will consist entirely of simplified Chinese characters that correspond to those comprising the primary domain name. The second variant will consist exclusively of traditional Chinese characters that correspond to those comprising the primary domain name.

The primary domain name together with the requested variants constitutes a bundle on which all operations are atomic. For example, if the registrant adds a name server to the primary domain name, all names in the bundle will be associated with that new name server.

The zone has no DNAME records, so the second paragraph strongly suggests parallel NS.

The .TEL TLD, intended as an online directory, does not allow registrants to enter arbitrary RR's in the zone. Nearly all names have NS records pointing to Telnic's own name servers. The A records all point to Telnic's own web server that shows directory information. NAPTR records provide the telephone number of registrants for whom they have one. Users can only directly provision MX records. Except that there are 16 domains, none IDNs, that point to random other name servers and mostly appear to be parked.

[4.17.](#) **TRAVEL**

The .TRAVEL TLD has no IDNs, and no rules or practices for them.

[4.18.](#) **XXX**

The .XXX TLD has no IDNs, and no rules or practices for them.

[5.](#) **Note about the references REMOVE BEFORE PUBLICATION**

Many of the references below may appear to be incomplete. This is due to bugs in the current version of XML2RFC. Consult the XML for

full names and URLs.

6. References

Levine & Hoffman

Expires April 02, 2013

[Page 7]

- [ASIACJK] ".ASIA CJK (Chinese Japanese Korean) IDN Policies", May 2011.
- [G1] "Guidelines for the Implementation of Internationalized Domain Names, Version 1.0", June 2003.
- [G3] "Guidelines for the Implementation of Internationalized Domain Names, Version 3.0", Sept 2011.
- [IANAIDN] "Repository of IDN Practices", .
- [ICANNAGREE] "ICANN Registry agreements", .
- [ICANNBIZ9] "Appendix 9 of ICANN .BIZ Registry agreement", Dec 2006.
- [ICANNCATS] "Appendix S of ICANN .CAT Registry agreement", Mar 2006.
- [ICANNINFO9] "Appendix 9 of ICANN .INFO Registry agreement", Dec 2006.
- [ICANNMOBIS] "Appendix S of ICANN .MOBI Registry agreement", Nov 2005.
- [ICANNNAME11] "Appendix 11 of ICANN .NAME Registry agreement", Mar 2011.
- [LEWIS03] Lewis, E., "Letter from Ed Lewis to Paul Twomey", Oct 2003.
- [MUSEUMIDN] "Internationalized Domain Names (IDN) in .museum - Policies and terms of use", Jan 2009.
- [PIRIDN] "Expanding Multi-Lingual Options in Domain Name Versatility", Jan 2009.
- [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, [RFC 1035](#), November 1987.
- [RFC3743] Konishi, K., Huang, K., Qian, H. and Y. Ko, "Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean", [RFC 3743](#), April 2004.
- [RFC5890] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework",

[RFC 5890](#), August 2010.

[RFC6672] Rose, S. and W. Wijngaards, "DNAME Redirection in the DNS", [RFC 6672](#), June 2012.

Levine & Hoffman

Expires April 02, 2013

[Page 8]

[TELPOLICY]

".TEL Policies", Jan 2009.

[TURNER03]

Turner, B., "Letter from Ben Turner to Paul Twomey", Nov 2003.

[TWOMEY03A]

Twomey, P., "Letter from Paul Twomey to Edward Viltz", Oct 2003.

[TWOMEY03]

Twomey, P., "Letter from Paul Twomey to Ram Mohan", Aug 2003.

[TWOMEY04A]

Twomey, P., "Letter from Paul Twomey to Richard Tindal", July 2004.

[TWOMEY04B]

Twomey, P., "Letter from Paul Twomey to Geir Rasmussen", Aug 2004.

[TWOMEY04]

Twomey, P., "Letter from Paul Twomey to Cary Karp", Jan 2004.

[VRSNADDL]

"Additional Logic", .

[VRSNCHAR]

"Character Variants", .

[VRSNLANG]

"Scripts and Languages", .

[VRSNRULES]

"Registration Rules", .

Authors' Addresses

John Levine
Taughannock Networks
PO Box 727
Trumansburg, NY 14886

Phone: +1 831 480 2300
Email: standards@taugh.com
URI: <http://jl.ly>

Paul Hoffman
VPN Consortium

Email: paul.hoffman@vpnc.org>

