IDN GLOSSARY

ENGLISH

A-label

The ASCII-compatible encoded (ACE) representation of an internationalized domain name, i.e. how it is transmitted internally within the DNS protocol. A-labels always commence with the prefix "xn--". Contrast with U-label.

ACE (ASCII Compatible Encoding)

ACE is a system for encoding Unicode so each character can be transmitted using only a limited set of ASCII characters (i.e. a-z, 0-9 and "-"). This is used because applications that use the DNS protocol may not reliably handle other values.

ASCII (American Standard Code for Information Interchange)

ASCII is a common numerical code for computers and other devices that work with text. Computers can only understand numbers, so an ASCII code is the numerical representation of a character such as 'a' or '@'. When mentioned in relation to domain names or strings, ASCII refers to the fact that before internationalization only the letters a-z, digits 0-9, and the hyphen "-", were allowed in domain names.

Character

For the purposes of discussing IDNs, a "character" can best be seen as the basic graphic unit of a writing system, which is a script plus a set of rules determining how it is used for representing a specific language. However, domain labels do not convey any intrinsic information about the language with which they are intended to be associated, although they do reveal the script on which they are based. This language dependency can unfortunately not be eliminated by restricting the definition to script because in several cases (see examples below) languages that share the same script differ in the way they regard its individual elements. The term character can therefore not be defined independently of the context in which it is used.

In phonetically based writing systems, a character is typically a letter or represents a syllable, and in ideographic systems (or alternatively, pictographic or logographic systems) a character may represent a concept or word.

The following examples are intended to illustrate that the definition of a character is at least two-fold, one being a linguistic base unit and the other is the associated code point.

U-label 酒 : Jiu; the Chinese word for 'alcoholic beverage'; Unicode code point is U+9152 (also referred to as: CJK UNIFIED IDEOGRAPH-9152); A-label is xn—jj4

U-label 北京 : the Chinese word for 'Beijing', Unicode codepoints are U+5300 U+4EAC; A-label is xn—1lq90i

U-label 東京 : Japanese word for 'Tokyo', the Unicode code points are U+6771 U+4EAC; A-label is xn—1lqs71d

U-label ایکوم; Farsi acronym for ICOM, Unicode code points are U+0627 U+06CC U+0643 U+0648 U+0645; A-label is xn—mgb0dgl27d.

Country-code Name Supporting Organization (ccNSO)

A component of ICANN's policy development forums (a "constituency") that is responsible for discussing and developing policy relating to how ccTLDs are delegated.

Country-code Top-Level Domain (ccTLD)

A class of top-level domains only assignable to represent countries listed in the ISO 3166-1 standard. At present these are two-letter codes like ".UK", ".DE" etc., however in the future it is expected there will be non-Latin equivalents also available. Much of the policy-making for individual country-code top-level domains is vested with a local sponsoring organization, as opposed to other top-level domains where ICANN sets the policy. It is a requirement that ccTLDs are operated within the country they are designated so appropriate local laws, governments etc. have a say in how the domain is run.

DNS (Domain Name System)   The DNS makes using the Internet easier by allowing a familiar string of letters (the "domain name") to be used instead of the arcane IP address. So instead of typing 207.151.159.3, you can type www.internic.net.

DNS Zone   A section of the Domain Name System name space. By default, the Root Zone contains all domain names, however in practice sections of this are delegated into smaller zones in a hierarchical fashion. For example, the ".COM" zone would refer to the portion of the DNS delegated that ends in ".COM".

DNSSEC   A technology that can be added to the Domain Name System to verify the authenticity of its data. This works by adding verifiable chains of trust that can be validated to the domain name system.

Domain Name   A unique identifier with a set of properties attached to it so that computers can perform conversions. A typical domain name is "icann.org". Most commonly the property attached is an IP address, like "208.77.188.103", so that computers can convert the domain name into an IP address. However the DNS is used for many other purposes. The domain name may also be a delegation, which transfers responsibility of all sub-domains within that domain to another entity.

Domain Name Label   A constituent part of a domain name. The labels of domain names are connected by dots. For example, "www.iana.org" contains three labels — "www", "iana" and "org". For internationalized domain names, the labels may be referred to as A-labels and U-labels.

Generic Top-Level Domains (gTLDs)   A class of top-level domains that are used for general purposes, where ICANN has a strong role in coordination (as opposed to country-code top-level domains, which are managed locally). For policy reasons, these are usually subdivided into sponsored top-level domains and unsponsored top-level domains.

IDN (Internationalized Domain Name)   IDNs are domain names that include characters used in the local representation of languages that are not written with the twenty-six letters of the basic Latin alphabet "a-z". An IDN can contain Latin letters with diacritical marks, as required by many European languages, or may consist of characters from non-Latin scripts such as Arabic or Chinese. Many languages also use other types of digits than the European "0-9". The basic Latin alphabet together with the European-Arabic digits are, for the purpose of domain names, termed "ASCII characters" (ASCII = American Standard Code for Information Interchange). These are also included in the broader range of "Unicode characters" that provides the basis for IDNs.

The "hostname rule" requires that all domain names of the type under consideration here are stored in the DNS using only the ASCII characters listed above, with the one further addition of the hyphen "-". The Unicode form of an IDN therefore requires special encoding before it is entered into the DNS.

The following terminology is used when distinguishing between these forms:

A domain name consists of a series of "labels" (separated by "dots"). The ASCII form of an IDN label is termed an "A-label". All operations defined in the DNS protocol use A-labels exclusively. The Unicode form, which a user expects to be displayed, is termed a "U-label". The difference may be illustrated with the Hindi word for "test" -- परीका -- appearing here as a U-label would (in the Devanagari script). A special form of "ASCII compatible encoding" (abbreviated ACE) is applied to this to produce the corresponding A-label: xn--11b5bs1di.

A domain name that only includes ASCII letters, digits, and hyphens is termed an "LDH label". Although the definitions of A-labels and LDH-labels overlap, a name consisting exclusively of LDH labels, such as "icann.org" is not an IDN.

IDN Practices Repository	A repository on IANA's website where top-level domain registries contribute the IDN tables they use. This allows other registries to re-use the tables if they wish.
IDN SLDs or IDN 2LDs	Usually a reference for domain names with local characters at the second level, while the top level remains in ASCII-only characters. For example: [παράδειγμα .test] ("example.test" in Greek).
IDN Table	An IDN Table is a table listing all those characters that a particular TLD registry supports. If one or more of these characters are considered a variant this is indicated next to that/those characters. It is also indicated which character a particular character is a variant to. The variant tables usually hold characters representing a specific language, or they can be characters from a specific script. Therefore the variant table is sometimes referred to as 'language variant table', language table', script table' or something similar.
IDN TLDs	Usually the short reference for internationalized top-level domains, thus allowing the entire domain name to be represented by local characters. For example: [실례.테스트] ("example.test" in Hangul).
IDNA (Internationalized Domain Names in Application)	IDNA is a protocol defined in RFC 3490 by the Internet Engineering Task Force (http://www.ietf.org) that makes it possible for applications to handle domain names with non-ASCII characters. IDNA converts domain name strings with non-ASCII characters to ASCII domain name labels that applications that use the DNS can accurately understand. Not all characters used in the world's languages will be available for use in domain names. Hence IDNA is not able to convert all such characters into ASCII labels.
Internet Assigned Numbers Authority (IANA)	A department of ICANN tasked with providing the functions described in a contract between ICANN and the US Government. The functions relate to ensuring globally-unique protocol parameter assignment, including management of the root of the Domain Name System and IP Address Space. ICANN staff within this department is often referred to as "IANA Staff".
Internet Coordination Policy (ICP)	A series of documents created by ICANN between 1999 and 2000 describing management procedures. Three such documents were published before the numbering system stopped being used. Subsequent ICANN publications have not been given ICP numbers.
Internet Engineering Steering Group (IESG)	The committee of area experts of the IETF's areas of work, that acts as its board of management.
Internet Engineering Task Force (IETF)	The key Internet standardization forum. The standards developed within the IETF are published as RFCs. IANA's protocol parameter registries are closely aligned with the work of the IETF.
IPv4	Internet Protocol version 4. Refers to the version of Internet protocol that supports 32-bit IP addresses. This allows for approximately 4 billion unique IP addresses, which is not enough to cope with projected Internet demand in the next 5-10 years. Therefore, a new protocol called IPv6 has been developed that increases the number of possible IP addresses substantially.

IPv6

Internet Protocol version 6. Refers to the version of Internet protocol that supports 128-bit IP addresses. This protocol is not yet widely deployed, but allows for orders-of-magnitude more IP addresses than the more common IPv4 protocol.

ISO

International Organization for Standardisation. An international organization comprised mostly of national standardization agencies.

Label

A label is an individual part of a domain name. Labels are usually shown separated by dots; for example, the domain name "example.com" is composed of two labels: "example", and "com".

Languages | Scripts | Alphabets

Languages are used by speech communities. Scripts are used to write down information in the various languages and this is done by using the corresponding alphabets or alternative writing systems.

LDH (Letter, Digit, Hyphen)

The hostname convention defined in RFC 952 (later modified by RFC 1123) was used by top-level domain Registries before internationalization. This meant that domain names could only practically contain the letters a-z, digits 0-9 and the hyphen "-". The term "LDH code points" refers to this subset. With the introduction of IDNs this rule is no longer relevant for all domain names although with the use of IDNA, what appears in the DNS remains LDH.

Local Internet Community

The community of Internet users within a country who benefit from the country's top-level domain. Country-code top-level domains are delegated to sponsoring organisations to operate domains in the best interests of this community, particularly by implementing policies the community has developed.

MIME Type

A formalised text string that identifies the type of a file that is included in the headers of an email or web transmission. IANA maintains the registry of MIME types.

Policy Development Process (PDP)

The formal policy creation process employed by ICANN and by a number of its constituencies.

Protocol

Any form of inter-computer communication that has been standardized to ensure computers can communicate to one another. Internet protocols are usually standardized in RFCs.

Punycode

Punycode is the LDH-compatible encoding algorithm described in Internet standard [RFC3492], and in use today. This is the method that is used to encode IDNs into sequences of LDH ASCII characters in order for applications using the Domain Name System (DNS) to understand and manage the names. The intention is that domain name registrants and users will never see this encoded form of a domain name. The sole purpose is for the DNS to be able to resolve for example a URL containing local characters. For examples see A-label under "IDN". The prefix in a Punycode A-label is always "xn--". Hence this prefix is recommended to be reserved by top-level domain Registries in order to avoid confusion when/if registrations of IDNs are introduced under the respective top level domain.

Registrant

The entity that has acquired the right to use an Internet resource. Usually this is via some form of revocable grant given by a registrar to list their registration in a registry.

Registrar

An entity that can act on requests from a registrant in making changes in a registry. Usually the registrar is the same entity that operates a registry, although for domain names this role is often split to allow for competition between multiple registrars who offer different levels of support. See also domain name registrar.

Registry

The authoritative record of registrations for a particular set of data. Most often used to refer to domain name registry, but all protocol parameters that IANA maintains are also registries.

Registry Operator

The entity that runs a registry.

Request for Comments (RFCs)

A series of Internet engineering documents describing Internet standards, as well as discussion papers, informational memorandums and best practices. Internet standards that are published in an RFC originate from the IETF. The RFC series is published by the RFC Editor.

Root

The most central (or all-encompassing) authority of any naming or numbering system. Usually used to refer to the domain name system root (see Root Zone). However, IANA is also the root for IP addresses, and other systems.

Root Servers

The authoritative name servers for the Root Zone. These are considered unlike regular name servers in part because they are generally the most critical and heavily-used name servers. They are also special as they are not easily replaced, as changes to them needs to be stored in every name server worldwide in a hints file.

Root Zone

The top of the domain name system hierarchy. The root zone contains all of the delegations for top-level domains, as well as the list of root servers, and is managed by IANA.

Script

A script is a collection of symbols used for writing a language. There are three basic kinds of script. One is the alphabetic (e.g. Arabic, Cyrillic, Latin) and its individual elements are termed "letters". A second is ideographic (e.g. Chinese), the elements of which are "ideographs". The third is termed a syllabary (e.g. Hangul) and its individual elements represent syllables. The writing systems of most languages use only one script but there are exceptions such as, for example, Japanese that uses four different scripts, representing all three of the categories listed here.

In order to be used in the computing environment, each element of a script needs to be numerically encoded. A collection of symbols numbered in this fashion is called a "character set". A character set may include more than one script (e.g. the "Universal Character Set", popularly known as Unicode), or it may be restricted to a single script (e.g. US-ASCII, which to be correct does not even cover the entire Latin script). A rigorous distinction must be made between scripts and character sets.

The only character set relevant to IDNA is Unicode. This assigns a numerical "code point" and a "character name" to every element of every script. The script-based policies that ICANN attaches to IDNs will operate on the names of the scripts that appear in Unicode character names, or on the blocks in the Unicode Code Chart that are similarly headed with script names. These script names are apparent at http://www.unicode.org/charts/.

For the purpose of the Fast Track Process, requesters must provide information about which script the strings in their request is represented in. From a practical standpoint the drop-down menu available for requesters, and hence facilitated in the Fast Track Online Request System is based on the ISO15924 list. From an evaluation standpoint, the validation of script and languages is defined in the Section 3.2 to the Fast Track Final Implementation Plan, as various methods for the requesters to select from. See http://icann.org/en/resources/idn/fast-track

It is important to note that characters in scripts which do not appear in the Unicode Code Chart are completely unavailable for inclusion in IDNs.

Sub-domain

A domain that resides within another domain. For example, "www.icann.org" is a sub-domain of "icann.org", and "icann.org" is a sub-domain of "org". Sub-domains are entrusted to other entities through a process of delegation.

The Unicode Consortium	A not-for-profit organization founded to develop, extend and promote use of the Unicode standard. For more information, please visit http://www.unicode.org.
Top-Level Domain (TLD)	The highest level of subdivisions with the domain name system. These domains, such as ".COM" and ".UK" are delegated from the DNS Root zone. They are generally divided into two distinct categories, generic top-level domains and country-code top-level domains.
U-label	The Unicode representation of an internationalized domain name, i.e. how it is shown to the end-user. Contrast with A-label.
Unicode	Unicode is a commonly used single encoding scheme that provides a unique number for each character across a wide variety of languages and scripts. The Unicode standard contains tables that list the "code points" (unique numbers) for each local character identified. These tables continue to expand as more and more characters are digitalized. In Unicode, characters are assigned codes that uniquely define every character in many of the scripts in the world. These "code points" are unique numbers for a character or some character aspect such as an accent mark or ligature. Unicode supports more than a million code points, which are written with a "U" followed by a plus sign and the unique number in hexadecimal notation; for example, the word "Hello" is written U+0048 U+0065 U+006C U+006C U+006F.
URL	An acronym for "Uniform Resource Locator", a string that describes the address of documents and other resources on the Internet. Defined by the IETF in RFC 2396, a URL is comprised of two parts separated by a colon (":"). The first part of the address indicates what protocol to use, e.g., http, ftp, etc., and the second part specifies the IP address or the domain name where the resource is located.
UTF-8	UTF-8 -bit Unicode Transformation Format is a system for encoding Unicode so each character can be transmitted using 8-bit numerical values. This is commonly used as 8-bit data transmission is prevalent on the Internet.
Variant	In the context of internationalized domain names, an alternative domain name that can be registered, or mean the same thing, because some of its characters can be registered in multiple different ways due to the way the language works. Depending on registry policy, variants may be registered together in one block called a variant bundle. For example, "internationalise" and "internationalize" may be considered variants in English.
Variant Bundle	A collection of multiple domain names that are grouped together because some of the characters are considered variants of the others.
Variant Table	A type of IDN table that describes the variants for a particular language or script. For example, a variant table may map Simplified Chinese characters to Traditional Chinese characters for the purpose of constructing a variant bundle.
Internationalized Registration Data Expert Working Group	
Plain ASCII	

Logogram

Unified ideograph

Blocked variant

Universal Acceptance

Trouble Ticket

stroke (o with)

IDN Radical

Arabic

Bengali

Cyrillic

Devanagari

Georgian

Greek

Gujarati

Gurmukhi

Han

Hangul

Hebrew

Hiragana,

Kannada

Katakana

Lao

Latin

Malayalam

Oriya

Sinhala

Tamil

Telugu

Thai