



Expedited Policy Development Process (EPDP) on Internationalized Domain Names (IDNs)

Satish Babu
Member, EPDP on IDNs



Overview

1. Common questions
2. Basic information on IDNs
3. The work done so far
4. The EPDP on IDNs: Rationale and scope
5. The Current Status



Common Questions

1. What are IDNs, IDN variants, PDP, EPDP and Universal Acceptance?
2. Why is the EPDP on IDNs important to end-user
3. The ICANN community has been working on IDNs since 2011. Why did it take 10 years to work on an IDN policy?
4. Why an EPDP and not a regular PDP?
5. How are IDNs delegated in the DNS?
6. What are IDN variants? Why are they important?
7. Will UA be considered in the EPDP?
8. What else will be considered in the EPDP?



The Basics

- Domain names are composed of labels separated by dots
- Each label is composed of the letters a-z, numerals 0-9, and the hyphen of the Latin character set (also called LDH). Each label can be up to 63 characters long
- It is not possible to use any other script, special characters or emojis to construct labels in the basic DNS
- This excludes numerous scripts/writing systems used by numerous languages around the world and precludes a truly multilingual Internet
- In order for an inclusive Internet, we need working domain names in multiple scripts (and URLs, email IDs and other artifacts using such domain names)



Definitions

- An Internationalized Domain Name is a domain name that has at least one non-Latin character in its labels
- There are three basic ways in which IDNs are implemented at present:
 - IDNs at the second level (for example **Bücher.com** or **भारत.org**) (since 2003)
 - IDN-based ccTLDs (**www.ไทย.ไทย**, **федеральнаяплощадка.рф**, **தளம்.பாராளுமன்றம்.இலங்கை**) (since 2010)
 - New IDN gTLDs (**도메인.삼성**, **nic.大众汽车**) (Since 2013)
- Because the DNS still resolves using the LDH, IDNs are internally converted to LDH through ASCII Compatible Encoding (ACE). The resulting “Punycode” looks like “**xn--bcher-kva.com**” or “**xn--hq1bm8jm9l.xn--cg4bki**”



Prior work

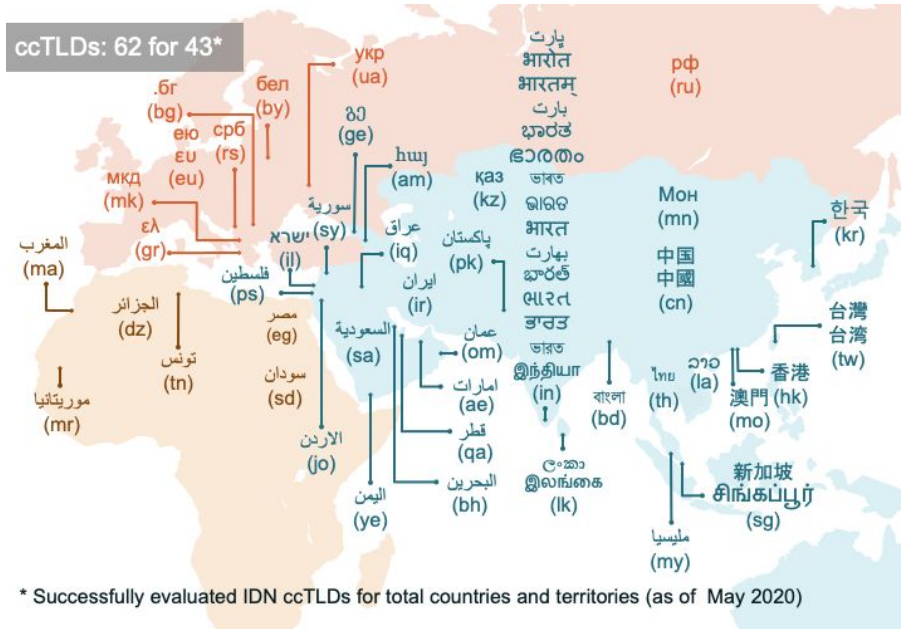
- These developments were made possible thanks to work done by the Unicode Consortium, IETF and the ICANN community & Board
- An important document that was helpful in the early adoption of IDNs was the IDN Implementation Guidelines which provided the mechanism for IDNs at the second level
- The 1.0 version was approved by the ICANN Board in 2003, after which Registries began to offer IDN-based domain names
- The Guidelines were amended 4 times. The fourth was completed in 2018, but as GNSO has asked for further review, the current version in effect is 3.0



ccTLD-based IDNs

- In 2009, the ICANN Board approved “Final Implementation Plan for IDN ccTLD Fast Track Process”
 - The implementation plan requires strings to be meaningful representation of a country/territory and supported by the government and the community
 - It also requires string to be reviewed by independent experts for technical considerations and string similarity
- The implementation plan has undergone multiple revisions for incorporating enhancements, with the current version being published in 2019

Status of IDN ccTLDs





IDN Generic Top Level Domains

- In 2007, while preparing for the next round of gTLDs, GNSO suggested several principles and recommendations including:
 - Principles
 - Some new generic top-level domains should be internationalised domain names (IDNs) subject to approval of IDNs in the root zone
 - The reasons for introducing new top-level domains include that there is demand from potential applicants for new top-level domains in both ASCII and IDN formats.
 - Recommendation
 - If an applicant offers an IDN service, then ICANN's IDN Guidelines must be followed



The first 4 IDN gTLDs

- شبكة (xn--ngbc5azd) – the Arabic word for “Web” or “Network”
- онлайн (xn--80asehdb) – Russian for “Online”
- сайт (xn--80aswg) – Russian for “Web site”
- 游戏 (xn--unup4y) – Chinese for “Game”



IDN Variants

- Language communities may use two or more labels as equivalent. For instance, in English we use “color” and “colour” interchangeably, or “æ” can also be written as “æ”
- Some languages (eg., Arabic) have large numbers of variant characters
- Some languages have multiple scripts. For instance, Chinese has traditional and simplified scripts. The same word looks different depending on the script, but the community considers both to be the same
- But to the computer, these are completely different strings. We have to take special action if we want both of them to behave the same way as domain names



IDN Variants (2)

- Variants can be a cause for user confusion
- If a label uses multiple variant codepoints (characters encoded in Unicode), then the number of label variants that need to be managed (ie., Allocated or Blocked) rises exponentially
- We have found it difficult to precisely define IDN variants and have so far side-stepped the issue for through temporary placeholders



IDN Variants for ccTLDs

- In the IDN ccTLD Fast-track process, the following was the placeholder mechanism:
 - Variant TLDs desired by the requester for delegation must be indicated by the requester
 - Desired variant TLDs will be allocated to the requester (if successfully evaluated) in order to be reserved to the entitled manager for potential future delegation in the DNS root zone
 - A list of non-desired variants will be generated based on the received IDN Tables and placed on a blocked list by ICANN
- The ccNSO community is expected to continue work on both defining variant ccTLDs and the mechanism of delegating them



IDN Variants for gTLDs

- The Applicant Guide Book of the 2012 round of new gTLDs had the following instructions:
 - The applicant may also declare any variant strings for the TLD in its application
 - However, no variant gTLD strings will be delegated through the New gTLD Program until variant management solutions are developed and implemented
 - Declaring variant strings is informative only and will not imply any right or claim to the declared variant strings
- When a variant delegation process is established, applicants may be required to submit additional information such as implementation details for the variant TLD management mechanism, and may need to participate in a subsequent evaluation process, which could contain additional fees and review steps.



Initial Work on IDN Variant TLDs

- The ICANN Board resolved in 2010 as follows:

No variants of gTLDs will be delegated through the New gTLD Program until appropriate variant management solutions are developed

- In 2011, ICANN launched the Variant Issues Project for five languages (Arabic, Chinese, Cyrillic, Devanagari, Greek, and Latin), which identified two main gaps



Gap 1: No definition of IDN Variant TLDs

- The solution was to define them using “Root Zone Language Generation Rules” Procedure (RZ-LGR Procedure), which was developed through community consultations
- On 11 Apr 2013, the ICANN Board resolved to implement the RZ-LGR procedure
- The procedure involves a Generation Panels and an Integration Panel
- A GP generally looks at one script (with exceptions such as the Neo-brahmi Panel with 9 scripts). In the case of the Chinese script, there are three GPs (Chinese, Korean and Japanese). Coordination amongst panels is also therefore required
- The IP is a single panel that integrates the output of GPs

Delegating IDNs

Generation Panels

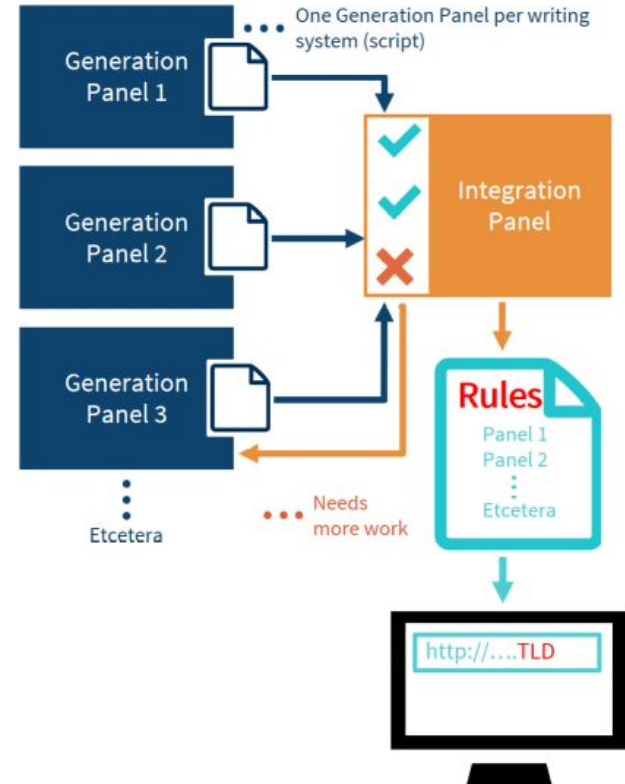
- Generate proposals for script specific LGRs, based on community expertise and linguistic, security and stability requirements

Integration Panel

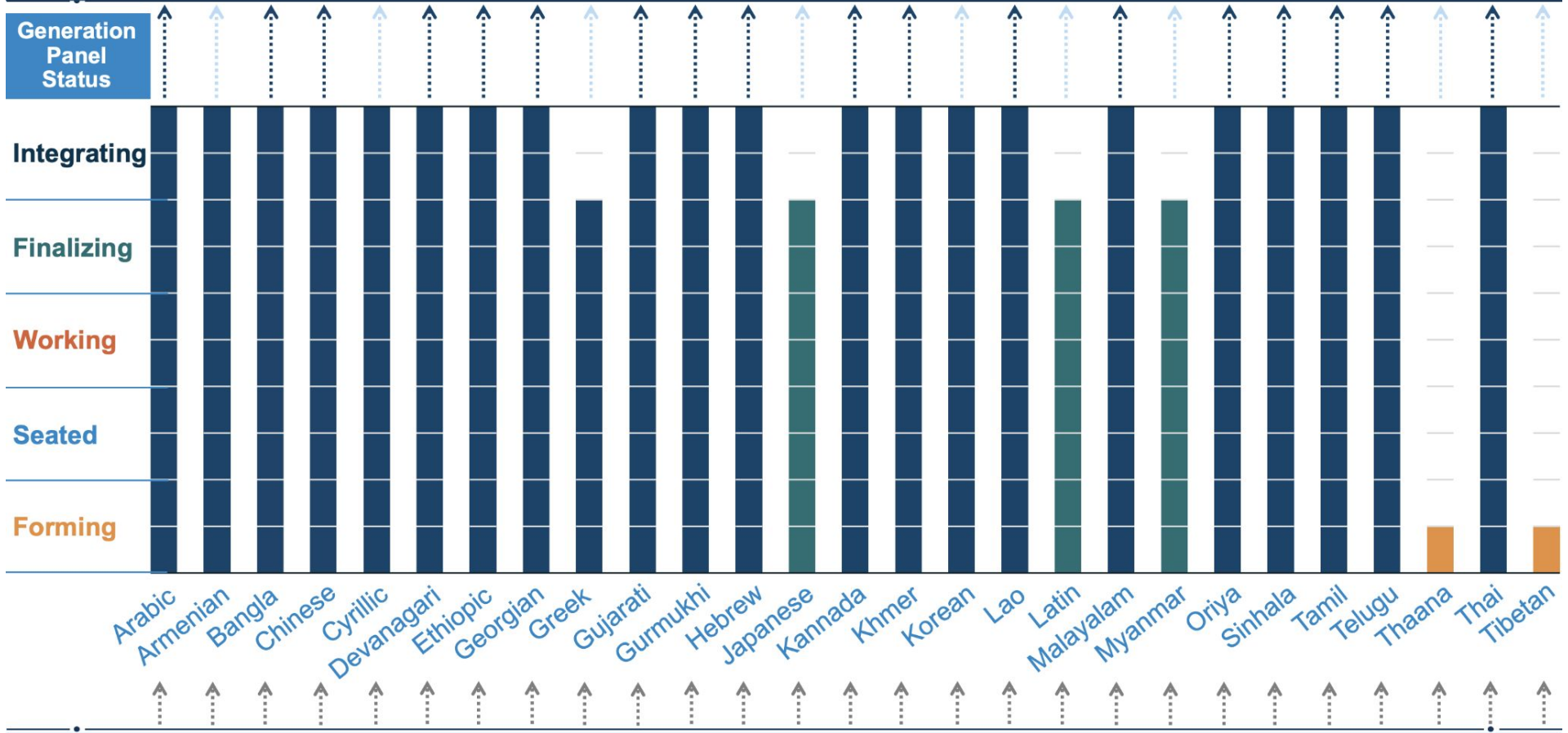
- Integrates them into common Root Zone LGR while minimizing the risk to Root Zone as shared resource

Label Generation Rules (LGR)

- Which labels are permissible?
- Which variant labels exist?
- Are there any more constraints?



Root Zone Label Generation Rules (RZ-LGR)



Maximal Starting Repertoire (MSR)



General Principles for TLDs

- TLDs are intended for “Unambiguous labels with good mnemonic value”
- Not intended to capture all facets of a writing system
- Should focus on modern, everyday use
- It’s OK not to support some conventions. For example, disallowing apostrophe does not support the ‘s ending for names of businesses
- Some limits necessary to reduce systemic risks



General Principles (2)

- The starting point for the delegation of a new script is the full set of Unicode characters of a script
- From the full list, a working set of characters is created, by dropping punctuation, joiners, historic characters, digits, religious/liturgic...anything that is not in widespread daily use
- This set is called the Maximal Starting Repertoire (MSR). The current version of MSR is 5.0
- Label Generation Rules (LGR) are formal specifications of how to validate a label. LGRs are used at registration time and not in the DNS
- Since DNS labels are mnemonics, LGRs do not insist on grammar or meanings of labels



Gap 2: No IDN TLD Variant management mechanism

- In March 2019, ICANN Org published a 6-part report titled “IDN Variant TLD Recommendations” that laid out 9 recommendations
 - a. RZ-LGR must be the only source for valid TLDs and their variant labels.
 - b. IDN variant TLDs must be allocated to the same entity (eg., **.aesop** and **.æsoṑ**)
 - c. Same labels under variants must be registered to the same entity (eg., **myshop.aesop** and **myshop.æsoṑ**)
 - d. Second-level variants also must be registered to the same entity (eg., **aeon.aesop**, **æon.aesop**, **aeon.æsoṑ**, **æon.æsoṑ**)
 - e. Existing policies and associated procedures for TLDs must be updated to accommodate the recommendations for IDN variant TLDs.
 - f. All remaining existing TLD policies must apply to IDN variant TLDs, unless otherwise identified



SubPro WG's work on IDNs

- The SubPro WGs submitted its final report in Feb 2021. The report touches on IDNs (p 114-118):
 - IDN New gTLDs to continue to be an integral part of New gTLD Program – Aff 25.1
 - Generation of TLDs & variant labels subject to compliance & checking – Rec 25.2
 - All RZ-LGRs: determine which label is blocked or allocatable
 - IDNA2008 (RFCs 5890-5895 or successor(s))
 - Where feasible, use of algorithmic checking – IG 26.10
 - Where RZ-LGR not yet integrated, allow application, processing but not contracting or delegation subject to further evaluation – IG 25.3



SubPro's IDN Work (2)

- Single character (U-label) gTLDs allowed for limited script/language combos – Rec 25.4
 - Ideograph or ideogram; Must not introduce risk of confusion from commonplace similarities
 - Conservative approach consistent with SAC052, Joint ccNSO-GNSO-IDN Workgroup (JIG) reports



GNSO IDN Scoping

- Based on ICANN Board resolutions, GNSO set up an IDN Scoping Team which submitted its report in Feb 2020. The report proposed two tracks of work:
 - Track 1: A working group that would be focused on IDN Guidelines v4.0 operational issues.
 - Track 2: An EPDP that would include:
 - How to define, manage, and coordinate IDN Variant TLDs
 - How the IDN Guidelines should be revised in the future
- In November 2020, GNSO set up a Charter Group to scope the EPDP, which produced its final report in May 2020



Detailed Scope of EPDP on IDNs

1. Consistent definition and technical utilization of RZ-LGR
2. “Same entity” at the top-level
3. “Same entity” at the second level
4. Adjustments in registry agreement, registry service, registry transition process, and other processes/procedures related to the domain name lifecycle
5. Adjustments to objection process, string similarity review, string contention resolution, reserved strings, and other policies and procedures
6. Adjustments in registration dispute resolution procedures and trademark protection mechanisms
7. Process to update the IDN Implementation Guidelines



Current Status

- The EPDP Team/Council has been seated and the the first meeting is scheduled for 11 Aug 2021
- The output of the EPDP would make it feasible to include IDN variant labels in the new gTLDs, and also recommend a mechanism for the periodic updating of the IDN Implementation Guidelines
- This would enhance user trust in the Internet and paves the way for a truly multilingual, inclusive Internet

Thank you!