



86

POLICY
FORUM



Internationalized Domain Names

How It Works

9 June 2026

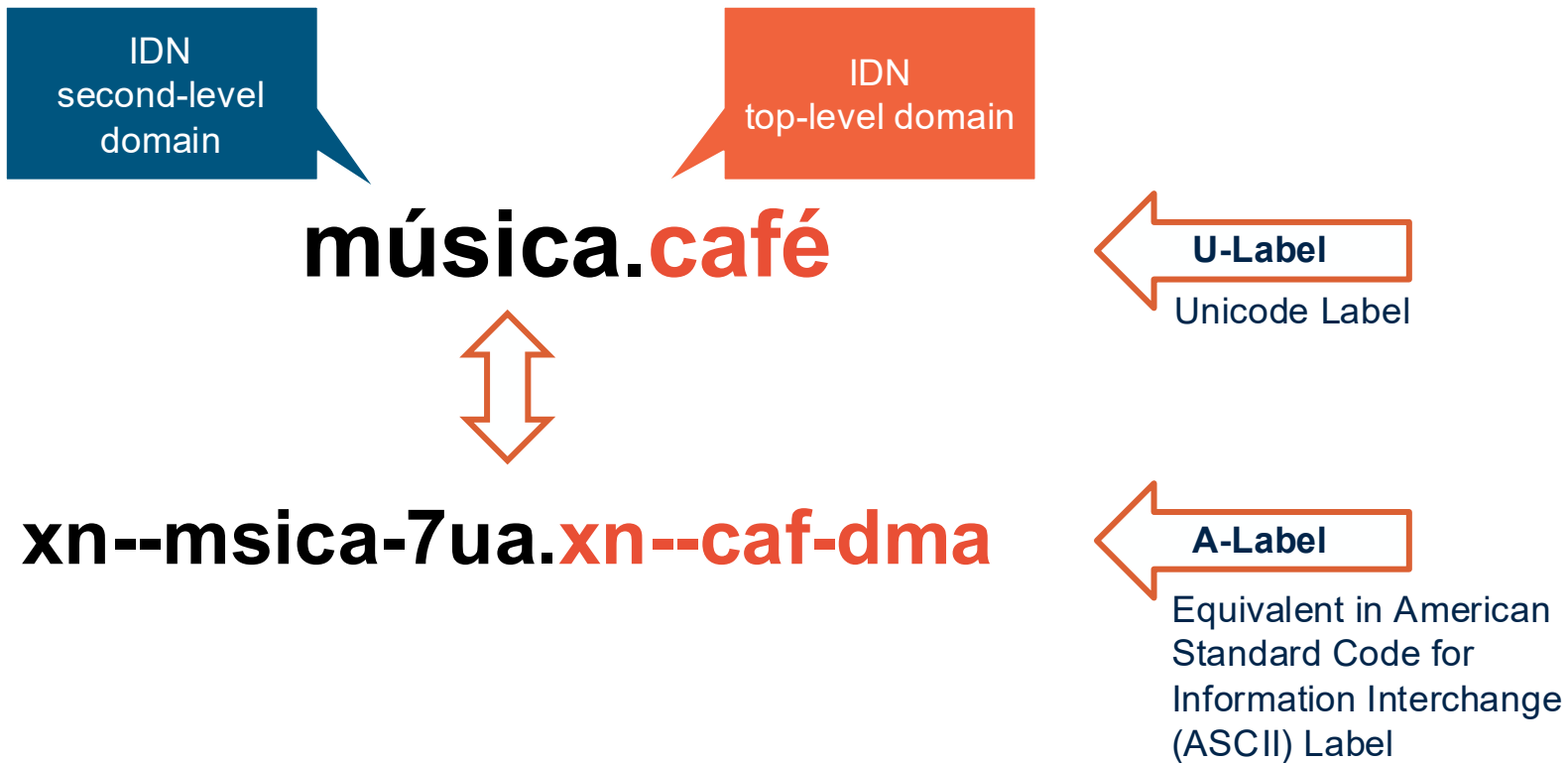
Agenda

- The process to enable Internationalized Domain Names (IDNs) in local languages and scripts
- IDNs Adoption in Europe
- IDN and Universal Acceptance (UA) Strategy and Case Study
- Q & A

The Process to Enable IDNs in Local Languages and Scripts

Pitinan Kooarmornpatana
IDN Program Director
ICANN

Internationalized Domain Name (IDN) Labels



IDNs for Applications (IDNA) 2008

- Internationalized Domain Names for Applications (IDNA) is a set of standards for handling Internationalized Domain Names (IDNs). Developed by Internet Engineering Task Force (IETF).
- The current version is IDNA 2008, consisting of:
 - Standard Track:***
 - [RFC 5890](#) IDNA: Definitions and Document Framework
 - [RFC 5891](#) IDNA: Protocol
 - [RFC 5892](#) The Unicode Code Points and IDNA
 - [RFC 5893](#) Right-to-Left Scripts for IDNA
 - Informational Track:***
 - [RFC 5894](#) Background, Explanation, and Rationale
 - [RFC 5895](#) Mapping Characters for IDNs in Applications (IDNA) 2008
- All IDNs must be IDNA 2008 compliant.

RFC 5892: The Unicode Code Points and IDNA

The value of the property (of a Unicode Code Point) is to be interpreted as follows.

PROTOCOL VALID (PVALID)*

Those that are allowed to be used in IDNs.

Example:

“é”

U+00E9

CONTEXTUAL RULE REQUIRED

Not to be used in labels unless specific context is met.

Example:

“l-l”

U+006C U+00B7 U+006C

(U+00B7 must be between 'l' (U+006C))

DISALLOWED

Those that should clearly not be included in IDNs.

Example:

“?”

U+003F
(Question Mark)

UNASSIGNED

Code points that are not designated in the Unicode Standard.

Example:

U+05C8 -
U+05CF

*A label consists only of PVALID code points does not imply that the label can be used in DNS.

Additional Requirements in IDNA2008

- **RFC 5892, Introduction section states:**
“...However, that a label consists only of code points that have this property value [PVALID] does not imply that the label can be used in DNS.”
- **RFC 5890, Section 2.3.2.3 states:**
*“Because of the diversity of characters that can be used in a U-label and the confusion they might cause, such restrictions [**“variant definitions and rules beyond those imposed by DNS or IDNA”**] are mandatory [emphasis added] for IDN registries and zones even though the particular restrictions are not part of these specifications...”*
- **Additional restrictions beyond IDNA are mandatory, therefore, the Label Generation Rules (LGRs)**

Label Generation Rules (LGRs)

Label Generation Rule (LGR)



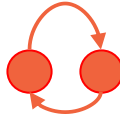
Metadata

Version, Date, Language, Description, Reference, etc.



Code Point Repertoire

Code Points, Code Point Sequences



Variants

Variant Code Points and Their Types



Rules

Contextual Rules, Whole Label Rules, Action Rules

Designing Principles

Starting from [RFC 6912](#) - Principles for Unicode Code Point Inclusion in Labels in the DNS:

- **Longevity** – stable across Unicode versions
- **Least Astonishment** – take into account the population using a code point
- **Contextual Safety** – sensitive to ways in which code point may be used in malicious ways
- **Conservatism** – any code point inclusion decision is as conservative as practicable
- **Inclusion** – default is excluded, then add code point which is safe based on usability and confusability
- **Simplicity** – rules determining use should be simple to understand
- **Predictability** – rules determining whether a code point is included are predictable for others to reach the same conclusion
- **Stability** – if permitted, taking it out very hard

Code Point Repertoire

What are code points to be used?

Code Point Repertoire - What is the Goal?

- **To create a mnemonic system for use in the Domain Name System (DNS)**
 - A mechanism to remember IP addresses
 - Must remain secure and stable in use – if DNS is confusing to users, then the motivation is not met
- **Not required to completely cover a language or a script**
 - May not form a word in a language
 - Not restricted to “correct” spellings
 - May not carry a meaning in the “lexical” sense

ASCII Domain Name Label

www.cafe-123.com

Forming ASCII Labels at Second-Level, Third Level

Use **LDH**

- Letters [a - z]
- Digits [0 - 9]
- Hyphen [H]

Label length = 63

Other constraints e.g. on Hyphen

Forming ASCII Label at Top-Level Domain

Use **only Letters**

- Letters [a - z]

Label length = 63

Domain Name Mnemonics in ASCII

- Top-Level Domain
 - Letters [a - z]
- Second-Level use LDH
 - Letters [a - z]
 - Digits [0 - 9]
 - Hyphen (H)

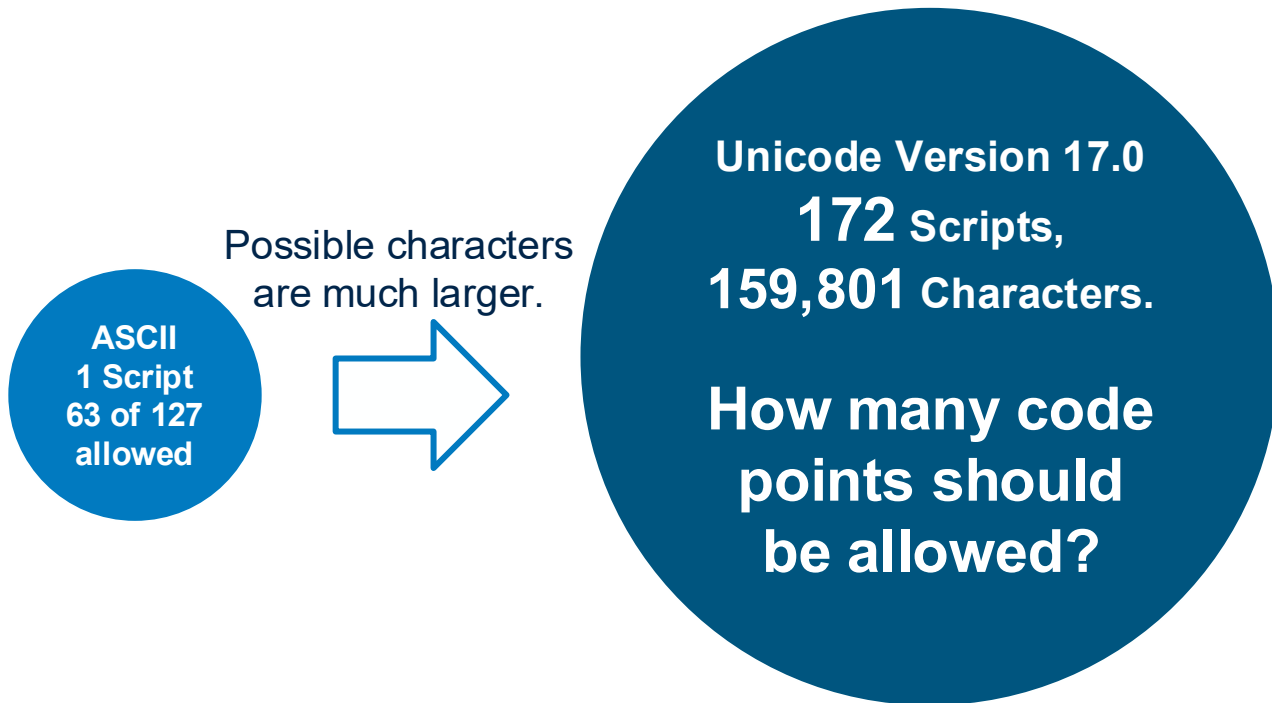
ASCII Printable Characters (Hex Code)

| | 2 | 3 | 4 | 5 | 6 | 7 |
|---|-------|---|---|---|---|-----|
| 0 | space | 0 | @ | P | ` | p |
| 1 | ! | 1 | A | Q | a | q |
| 2 | " | 2 | B | R | b | r |
| 3 | # | 3 | C | S | c | s |
| 4 | \$ | 4 | D | T | d | t |
| 5 | % | 5 | E | U | e | u |
| 6 | & | 6 | F | V | f | v |
| 7 | ' | 7 | G | W | g | w |
| 8 | (| 8 | H | X | h | x |
| 9 |) | 9 | I | Y | i | y |
| A | * | : | J | Z | j | z |
| B | + | ; | K | [| k | { |
| C | , | < | L | \ | l | |
| D | - | = | M |] | m | } |
| E | . | > | N | ^ | n | ~ |
| F | / | ? | O | _ | o | del |

Repertoire for IDN Mnemonics

- Unicode Standards includes code charts for all encoded scripts. See the full Unicode Code Chart at <https://www.unicode.org/charts>
- A subset of code points suitable to be used in a domain name are defined by the script community

ASCII vs Unicode



Building the Code Point Repertoire

Step 1: Review relevant script Unicode code chart(s)

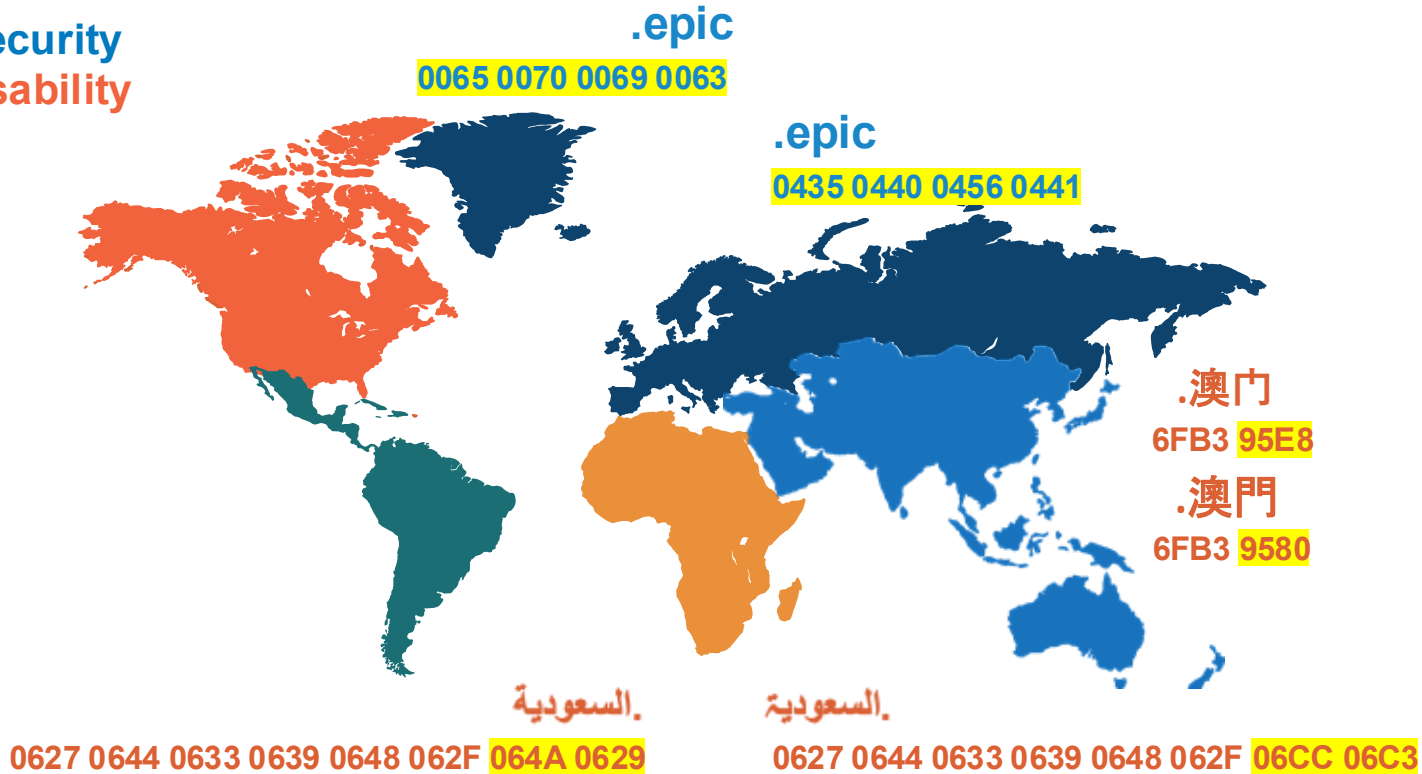
Step 2: Before including a code point, ask the following questions:

- Is it PVALID?
- Is it used with the language or script in scope?
- Is it suitable in identifiers?
 - Is it in widespread modern use?
 - Is it not technical / religious / limited use only?
 - Is it not really a punctuation / symbol?
 - Is it really necessary for representing identifiers?
- Is the Unicode encoding of the code point stable?
 - Are there any rendering issues?
- How accessible would a TLD containing that code point be?
 - Are there input/keyboard concerns?
- Is it in tension with any of the RFC6912 Principles in any way?

***Are there any code points considered “same”
by the language or script community?***

Understanding Variants

- Security
- Usability





Variants - What is the Goal?

- **Make the mnemonic system to minimize user confusion**
 - Defining code point or code point sequences causing two (or more) labels functionally “the same” within script, and cross-script.
- **Conservatism requires:**
 - maximize “blocked” variants
 - minimize “allocatable” variants
- **Note: Successfully defining variant rules for an LGR is not trivial**

Which Scripts Have Variant Code Points?

- Arabic
- Armenian
- Bengali
- Cyrillic
- Devanagari
- Ethiopic
- Georgian
- Greek
- Gujarati
- Gurmukhi
- Han
- Hebrew
- Japanese
- Kannada
- Khmer
- Korean
- Lao
- Latin
- Malayalam
- Myanmar
- Oriya
- Sinhala
- Tamil
- Telugu
- Thaana
- Thai

| | |
|--|------------------------|
|  | Variant code points |
|  | No variant code points |

Examples of Cross-Script Variants

| Cyrillic | Greek | Latin |
|---|---|---|
| <p>а U+0430 CYRILLIC SMALL LETTER A</p> | <p>α U+03B1 GREEK SMALL LETTER ALPHA</p> | <p>a U+0061 LATIN SMALL LETTER A</p> |
| <p>о U+043E CYRILLIC SMALL LETTER O</p> | <p>ο U+03BF GREEK SMALL LETTER OMICRON</p> | <p>o U+006F LATIN SMALL LETTER O</p> |
| <p>р U+0440 CYRILLIC SMALL LETTER ER</p> | <p>ρ U+03C1 GREEK SMALL LETTER RHO</p> | <p>p U+0070 LATIN SMALL LETTER P</p> |

Defining Variants - Questions to Ask

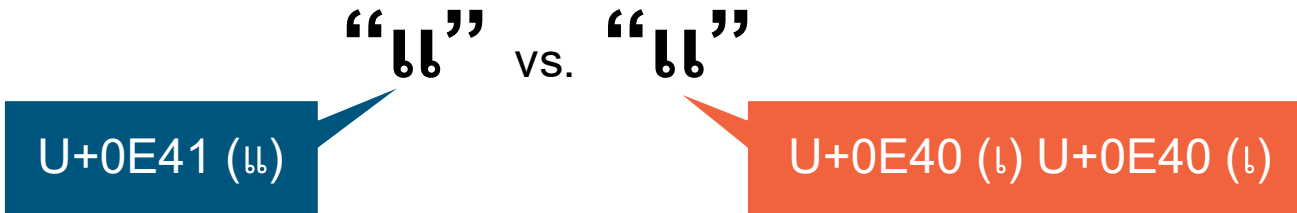
- Would a reasonable person with native knowledge of the script consider a pair of code points interchangeable?
- Would such a person be unable to determine which of these interchangeable code points was used by appearance?
- Is there an alternative representation?
- Should any of the variants of this code point be contingent on context?
- Are any variants contemplated that are in tension with any of the Principles?
- Any cross-script variants?

Are there any restrictions required to ensure the language or script is used in a secure and stable manner in the DNS?

Rules - What is the Goal?

- **Reducing label space, preventing labels which should not be possible for various reasons**
 - Not licensed by the script (but not spelling rules)
 - Cause security issues
 - Cause usability constraints
 - Other?
- **Reducing allocatable labels by making them blocked in certain cases**
- **Adding contextual constraints for code points or their sequences**

Example of Risk Mitigation (Thai Script)



- Double U+0E40 (Thai Character Sara E) look like a U+0E41 (Thai Character Sara Ae)
- Thai script LGR includes a rule to prevent U+0E40 from follow another U+0E40

Example of Risk Mitigation (Lao Script)

“U+0EAE (ຮ) U+0EC9(ຶ) U+0EC9(ຶ)”

Consonant

Tone Mark

Tone Mark



Chrome on Android 16



Chrome on iOS 26.3

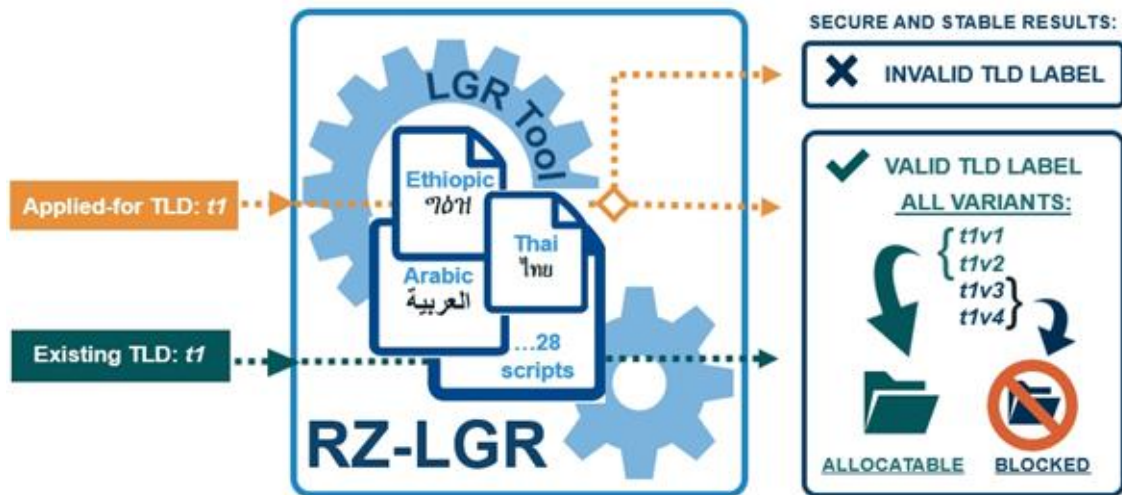
- The Lao script expects none or one tone mark on a consonant.
- More than one tone mark on a consonant is rendered inconsistently.
- Laos script LGR includes a rule to restrict one tone mark on a consonant.

LGR Tool

- The LGR Tool, provided by ICANN org, helps with developing and managing an LGR, validating labels against a LGR, and reviewing IDN tables.
 - To use the tool, please visit: [*lgrtool.icann.org*](https://lgrtool.icann.org).
It is available for all users with an ICANN account.
 - User Guides are available on the [***LGR Tool***](#) [*webpage*](#).
- The LGR Tool is an open source application. Parties interested in integrating this functionality into their own systems can find the source code on github: [lgr-core](#), [lgr-django](#), [munidata](#), [picu](#).

Usage of LGR

- LGR in machine-readable format can be processed by the LGR Tool to validate a label and generate or validate their variant labels
- For example: Root Zone Label Generation Rules



Example: Latin Label Validation

Based on the [Latin script Root-Zone LGR](#)

Test Label: “hii”

✓ VALID

| U-label | A-label | Disposition | Code point sequence |
|---------|---------|-------------|----------------------------------|
| hii | hii | valid | U+0068 (h) U+0069 (i) U+0069 (i) |

Latin Label: “hi!”

✗ INVALID

| U-label | A-label | Disposition | Code point sequence |
|---------|------------|-------------|----------------------------------|
| hi! | xn--hi-01a | invalid | U+0068 (h) U+0069 (i) U+01C3 (!) |

The code point U+01C3 (!) Latin Letter Retroflex Click is not in the Latin script LGR repertoire.

Example: Latin Label Validation

Required rule: The Latin middle dot (·) U+00B7 must be between 'l' (U+006C).

Test Label: “col·legi”

✓ **VALID**

| U-label | A-label | Disposition | Code point sequence |
|----------|-----------------|-------------|--|
| col·legi | xn--collegi-xma | valid | U+0063 (c) U+006F (o) U+006C (l) U+00B7 (·) U+006C (l) U+0065 (e) U+0067 (g) U+0069 (i) |

Latin Label: “col·com”

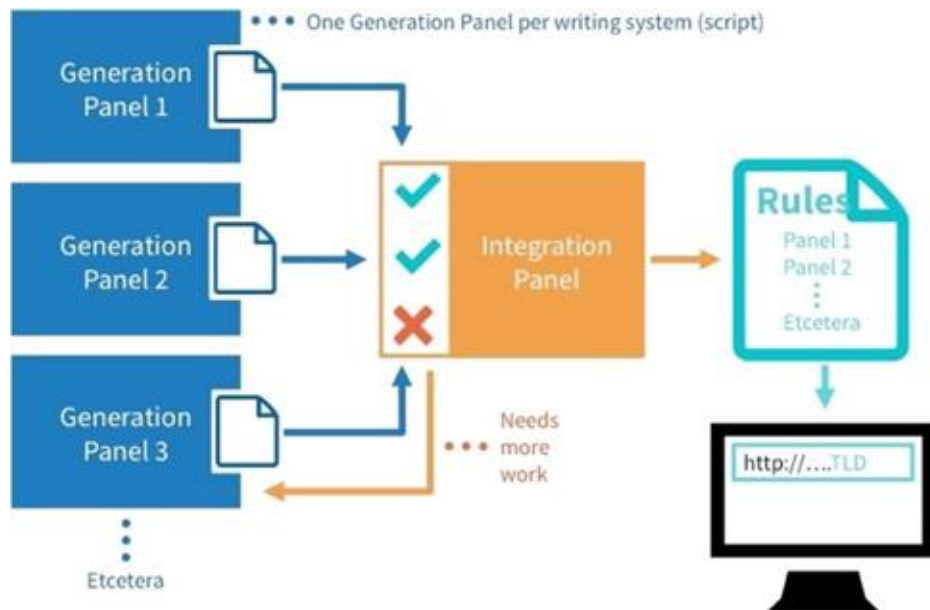
✗ **INVALID**

| U-label | A-label | Disposition | Code point sequence |
|---------|---------|-------------|--|
| - | - | invalid | U+0063 (c) U+006F (o) U+006C (l) U+00B7 (·) U+0063 (c) U+006F (o) U+006D (m) |

Need for Community-Driven Development of Rules for IDNs

Root Zone Label Generation Rules (RZ-LGR)

- RZ-LGR development process includes input from script communities
- RZ-LGR version 6
 - 27 scripts
 - supporting 387+ languages
- 11 years of work, 18 Generation Panels (GPs) from 45 countries
- 270+ script community volunteers
10,000+ volunteer hours



- RZ-LGR version 6 (RZ-LGR-6) includes 27 scripts
 - Arabic, Armenian, Bangla, Chinese (Han), Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Hebrew, Japanese (Hiragana, Katakana, and Kanji [Han]), Kannada, Khmer, Korean (Hangul and Hanja [Han]), Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, Thaana, and Thai

Reference LGRs for Second-Level (Ref. LGR)

- ICANN org developed second level reference LGRs and available RZ-LGR following consultations with the script communities, which were finalized after Public Comment proceedings.
- TLD registry operators can use reference LGRs when they develop rules for IDN registration under their TLDs.
 - ICANN org uses reference LGRs to review the registration rules (called IDN Table) under gTLDs.
 - This contributes to consistency and transparency of the review as requested by the Registries Stakeholder Group (RySG).
- Reference LGR can be used to develop an IDN Table which is required for submitting an IDN ccTLD Request via the IDN ccTLD Fast Track Process.

Reference LGRs for Second-Level (Ref. LGR)

ICANN publishes reference LGRs for 60 languages and scripts:

- **32 language-based reference LGRs:** Arabic, Belarusian, Bosnian (Cyrillic), Bosnian (Latin), Bulgarian, Chinese, Danish, English, Finnish, French, German, Hebrew, Hindi, Hungarian, Icelandic, Inuktitut, Italian, Japanese, Korean, Latvian, Lithuanian, Macedonian, Montenegrin, Norwegian, Polish, Portuguese, Russian, Serbian, Spanish, Swedish, Thai, and Ukrainian.
- **28 script-based reference LGRs:** Arabic, Armenian, Balinese, Bangla (Bengali), Chinese, Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Hebrew, Japanese (Hiragana, Katakana, Kanji [Han]), Kannada, Khmer, Korean (Hangul, Hanja [Han]), Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, Thaana, and Thai.

Public Comment: Additional Ref. LGRs

ICANN publishes additional reference LGRs for public comment:

<https://www.icann.org/en/public-comment/proceeding/additional-reference-label-generation-rules-and-related-updates-12-05-2026>

- **2 new reference LGRs:** Javanese script, Unified Canadian Aboriginal Syllabics (UCAS) script
- **6 updated reference LGRs:** Arabic script, Inuktitut language, Japanese language, Japanese script, Latin script, Latin script (Full Variant)

Closes for Submissions 23 June 2026, 23:59 UTC

Collaboration on Developing the Ref. LGR

Voluntary expertise provided by a community working group

- Linguistic expertise to develop a proposal that define repertoire, variants, rules.
- Word list or corpus for a language or script.
- Review and test the LGR.

Support provided by ICANN org

- Technical support on developing an LGR in RFC7940 format, based on input from the community.
- Taking an LGR through the public comment process before finalizing and publishing an LGR.



How to Get Involved

- If you'd like to develop a reference LGR for your language or script, please contact IDNProgram@icann.org
- Information on IDNs, please visit icann.org/idn

IDN Adoption in Europe

Regina Filipová Fuchsová
Industry Relations Manager
EURid, registry of .eu, .eu, and .eu

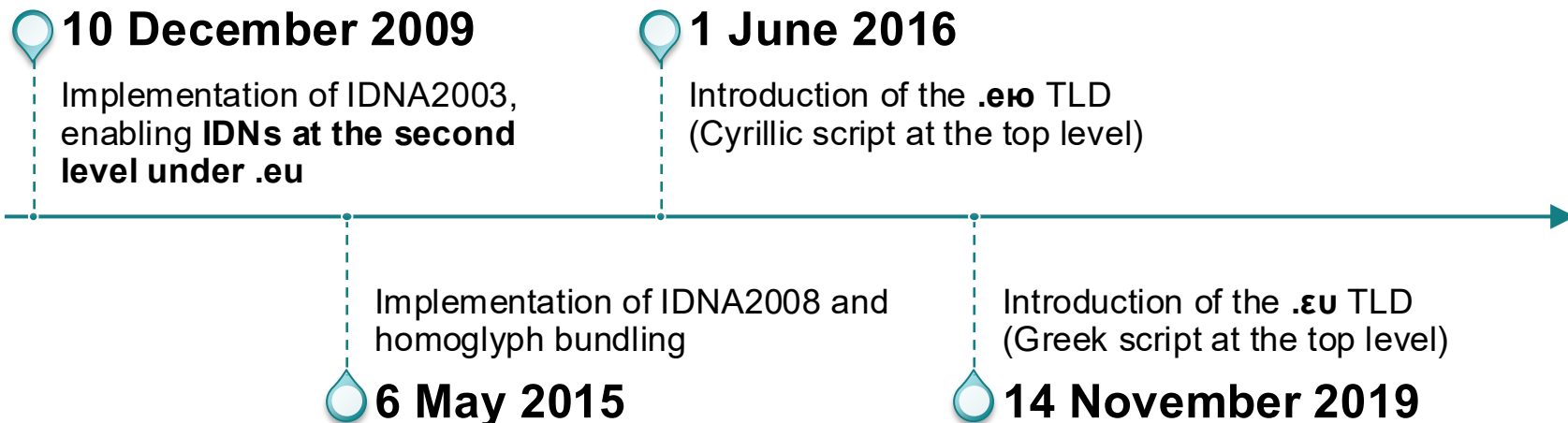
Accessibility and Multilingualism online

Since 2005 entrusted to provide registry services in all EU languages, by appointment of the EC.

Part of EURid' s strategy:

- Greek/Cyrillic variant of .eu and special characters
- Partnerships on IDNs/UA topics
- Active involvement in the ccNSO Universal Acceptance Committee (UAC)
- idnworldreport.eu

Technical Implementations of IDNs @EURid

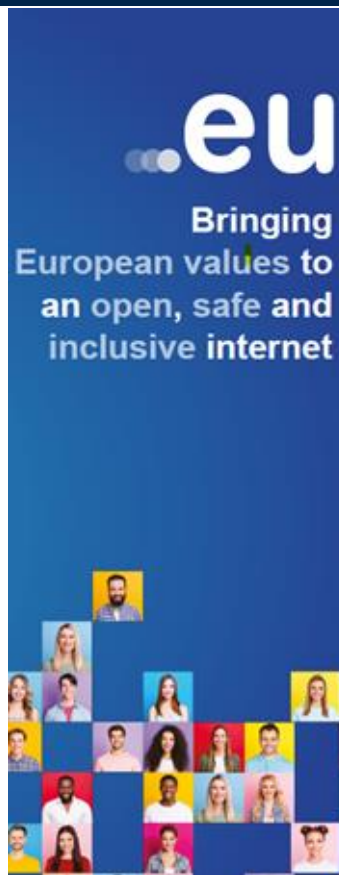


EURid: ccTLD Manager of .eu, .eu and .eu

3.8M registrations

0.9% IDNs (37.3k)

- 29.8k IDNs under .eu
- 2.1k under .eu
- 1.2k under .eu



Aggregated Data on the Use of IDNs at EURid (end 2025)

70% no content or just parking pages
30% some kind of user generated content
32.6k domains crawled



IDNWorldReport.eu

Top-Level Domain Names



- A long-standing **research project** of EURid, UNESCO and other partners
- A resource aimed at **enhancing linguistic diversity online** through advocacy of Internationalized Domain Names (IDNs)

.eu
POWERED BY EURid

New Edition Just Published

Provides articles, research and data on the development and deployment of **IDNs** across TLDs.

Recent global
registration
trends

Technical
readiness

Sentiment
analysis across
ccTLDs and
gTLDs

Complemented
by articles and
case studies

Registration data of 374 TLDs and a sentiment survey of 43 registries from across Asia, Europe, and Latin America.

Key Findings of the 2026 Edition



Market still
concentrated and
declining



Technical
readiness is
improving



Awareness
remains the
biggest barrier

The Global IDN Market: Registrations

- Data set of 71 ccTLDs and 303 gTLDs:
 - at least **4.3 million IDN registrations** (a slight decrease from 4.4 million in 2025)
 - of which roughly 71% are under a ccTLD or IDN ccTLD
- The top 10 TLDs: over 3.2 million domains (roughly 76%)
- The top five: .pф (779K) – positive growth, .com (733K), .de (618K), .cn (537K), and .中国 (185K).
 - .com and .net are the only two gTLDs represented in the Top 10

An overall global volume of registrations - limitations on the number of registries that supply data

Importance for Local Communities

.cat

à; ç; è é; í; ï; l-l; ò; ó; ú; ü

- Equivalent without special characters on request

Challenges:

- Awareness of end customers and registrars
- Improvements in UA: some email clients do not accept IDNs or offer bad UX (showing Punycode)
- Ambassadors: more visibility to domains using IDNs in different platforms

Highlight advantages:

- Get your domain in your language / IDNs shows your pride (when necessary) / Do not change the name of your project to fit / Be different and related to your culture

.fo

á, í, ó, ú, ý, æ, and ø

Advertising campaign upon launch aimed at Faroese citizens through various media, including national television and YouTube advertisements

.fo reaffirming cultural heritage and ensuring inclusion on the global digital landscape

Challenges:

- Business have to maintain both IDNs and non-IDNs
- IDN characters in email addresses (Punycode)

References to IDNs and UA in Internet Governance

“Users around the world can **navigate the Internet entirely in local languages**”

IDNs are necessary for multilingual Internet

IDNs, UA and WSIS:

- Shared goal: building an inclusive, multilingual, and globally accessible internet
- Tool that implements WSIS principles by making the internet linguistically inclusive, culturally diverse, and accessible to all

How to attract the next billion of Internet users – concepts of linguistic accessibility, digital inclusion, nobody left behind, link to SDGs

What's Next?

Further cooperation and synergies among partners (survey)

In-house improvements in UA

Further work within the ccTLD community - ccNSO Universal Acceptance Committee (library, mailing list, membership)

IDN and Universal Acceptance (UA) Strategy and Case Study

Dejan Đukić

.rs

How It Works: IDN?

IDN and Universal Acceptance strategy
and case study



Serbian language and linguistic diversity in Serbia

The Serbian language is one of the few languages in the world that uses two scripts: Cyrillic and Latin. This feature is called digraphia, which denotes the use of two writing systems for a single language.

In the Republic of Serbia, the Serbian language is in official use. The Cyrillic script is in official use, while the Latin script is used in the manner prescribed by law.

In areas of the Republic of Serbia where members of national minorities live, the languages and scripts of national minorities are in official use alongside the Serbian language.

Serbian ccTLDs – Serbian IDNs




- Since 2008
- Since 2018 supports on the second level Latin alphabets used by National minorities in Serbia: Albanian, Croatian, Czech, German, Hungarian, Montenegrin, Romanian, Romani, Slovak, Slovenian, Vlach (Aromanian)



- Since 2012
- Since 2019 supports Cyrillic alphabets used by National minorities in Serbia: Bulgarian, Macedonian, Montenegrin, Russian, Rusyn (Ruthenian), Ukrainian, Vlach (Aromanian)





**Promoting Serbian
IDNs by creating
content that creates
content**

Continuous project repeated annually to celebrate .cpb birthday by giving away free Serbian Cyrillic typefaces.

Engages academia, artists, designers, developers, but also media and broader public ...

Areal RNIDS
Ареал РНИДС

Etar RNIDS
Етар РНИДС



Orto RNIDS
Орто РНИДС

Manument RNIDS
Манумент РНИДС

Lingva RNIDS
Лингва РНИДС

Almono RNIDS
Алмоно РНИДС

Neo Geo RNIDS
Нео Гео РНИДС

*Also available in Latin script



RNIDS Serbian Cyrillic Fonts in Motion

- Websites of governmental bodies and private businesses
- Social media posts and digital publications of small businesses and major brands
- Brand books and brand guidelines
- Hackathons project assignments
- Publishing and children's literature



ИМА НА Бирлици ХАКАТОН

IDN themed hackathons

*First touch and UA related
experience for many young
delelopers and software
engineers*



Useful content provided, however,
more USE CASES needed.

Thank you for your attention.

Dejan Đukić

dejan.djukic@rnids.rs

www.rnids.rs
рнидс.срб

www.domen.rs
домен.срб

-  icann.org
-  [@icann](https://twitter.com/icann)
-  facebook.com/icannorg
-  youtube.com/icannnews
-  flickr.com/icann
-  linkedin.com/company/icann
-  instagram.com/icannorg