

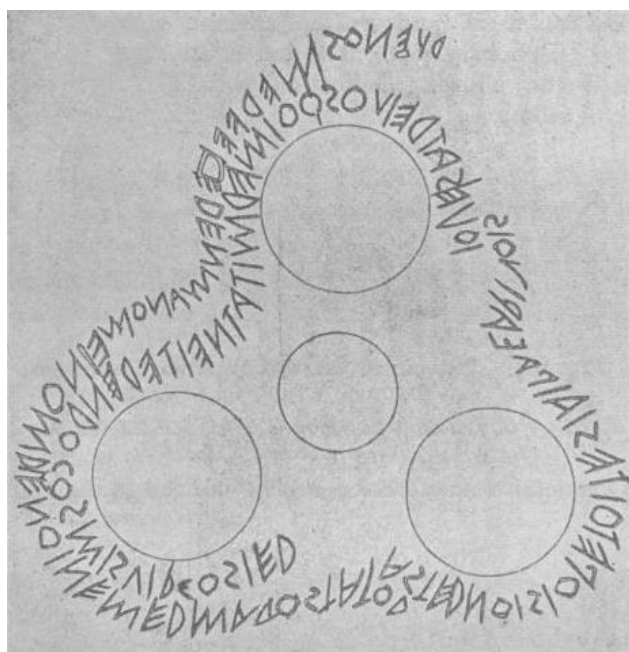
Proposal for Generation Panel for Latin Script Label Generation Ruleset for the Root Zone. Ed. C. Dillon. Version 6 (16 March 2016)

1. General Information

The Latin script¹ is derived from the Greek alphabet², as is the Cyrillic script. The Greek alphabet is in turn derived from the Phoenician alphabet which dates back to the mid-11th century BC and is itself based on older scripts. This explains why Latin, Cyrillic and Greek share some letters.

The Latin alphabet originated in Italy in the 7th Century BC. The original letters were: A, B, C, D, E, F, Z, H, I, K, L, M, N, O, P, Q, R, S, T, V and X. There were only upper case letters.

G developed from C and J from I. V and U split and a ligature of VV became W. Languages added new letters, for example þ (thorn) for Scandinavian languages, borrowed from the runic alphabet. Letters were often combined to form ligatures, (for example, æ from a and e in Danish and Norwegian) or ß (from Gothic s and z, in German). The current basic set is: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y and Z.



The Latin script is alphabetic – there are letters for both consonants and vowels. Some languages, such as Esperanto, use it phonemically, so that sounds are represented in a systematic way; other languages, such as English, use it so that other aspects, such as etymology, are represented. For example, the spelling of *night* connects it with German *Nacht*, although *gh* is no longer pronounced.

Letters of the Latin script now exist in upper and lower case forms. There may be little visual similarity between a letter's upper and lower case forms, for example, A and a.

The Duenos Inscription, 6th Century B.C.

The Latin script is almost always written left-to-right.

Letter shapes may be considerably different depending on the language and whether the script is handwritten or printed.

¹ *Script* is used here to indicate the whole writing system including basic letters, ligatures and diacritics. See also ISO 15924.

² *Alphabet* is used to refer to the basic set of letters, as used, for example, in a dictionary.

There are many different writing styles. Until the 1940s, for example, German was commonly written in Gothic (or blackletter) script ("Fraktur"). Sütterlin was a common form:

Leibnizmilch wird im
 form der deutschen
 Schrift oder Sütterlin-
 Schrift bezeichnet. Die
 Schrift wird durch, daß
 die Sütterlin-Schrift die
 neue form der deutschen
 Schrift ist, man
 kann um bezeichnen
 ist. Jedem ist die
 Schriftung anzusehen,
 man es gibt die deutsche
 Schrift schon von
 der 18. Litteratur
 lin.

Normally letters appear separately when printed and joined together when written by hand. However, some printed fonts join the letters together and many people have individual preferences for writing at least some letters separately in their handwriting.

Spaces are almost always used to separate words. The hyphen (-) is used in many languages to separate elements that belong together in some way, for example, parts of a compound word or to indicate that a word has been truncated, for example, at the end of a line.

Sample of printed Fraktur by -donald-

*Ewel zaman içinde,
 Kalbur saman içinde,
 Daveler tellal iken,
 Pireler kamal iken,
 Ben de babamın beşliğini
 Tüngen münğün sallan iken...*

Sample of Turkish handwriting. Note how T and i are not joined at the start of the last line.

Diacritics also came to be used to modify letters in many languages. These may appear anywhere around, most commonly above (é), below (ç), or through (ø) a letter. Several diacritics may attach to the same letter; Vietnamese ơ, for example, has a hook on the right and a dot below.

Some languages consider letter + diacritic as one letter. Norwegian (both Bokmål and Nynorsk varieties), for example, lists these three letters at the end of its alphabet: Æ, Ø and Å.

Diacritics may perform different roles depending on the language:

- For example, in French the acute accent over e (é) is used to indicate a closed e sound, for example, café.

- In Spanish, however, the same diacritic is used to indicate cases where the stress does not fall on the penultimate syllable, for example, dieciséis ‘sixteen’.
- In Vietnamese, the same diacritic would indicate a high rising tone.

1.1 Latin Script as Represented in Unicode

As represented in Unicode, the Latin script has some identical glyphs³, for example, 0259 ə (schwa) and 01DD ə (turned e). The following glyphs belong to both the Latin and Cyrillic scripts: a, b, e, s, i, j, k, m, o, p, c, t, y, and x (non-exhaustive list). Here only lower case glyphs are considered, as upper case ones may not be used in IDNs.

A letter with two diacritics, for example, ç, may be typically represented in several ways in Unicode – as a pre-composed form (U+1E09), or as the letter and the first diacritic with the second added (U+0107 ć + U+0327 COMBINING CEDILLA), or with the letter and the second diacritic with the second diacritic added (U+00E7 ç + U+0301 COMBINING ACUTE ACCENT).

Montenegrin has ligatures lj and nj which do not exist as pre-composed code points in Unicode 6.3 (on which IDNA 2008 is based). It is likely that there are similar cases in African languages.

It is possible that combining marks may be required for some languages in widespread modern use.

1.2 Target Script for the Proposed Generation Panel

The Latin script has the following specifications:

ISO 15924 code: Latn

ISO 15924 no.: 215

English Name: Latin

Note that the Gaelic and Fraktur variants of Latin have their own ISO 15924 codes and numbers (Latg 216 and Latf 217 respectively), and so do not fall within the remit of the Latin Generation Panel (LGP).

The complete set of code points in the Latin script fall in the following Unicode ranges:

| | |
|---------------------------------|-----------------|
| Controls and Basic Latin | U+0061 – U+007A |
| Controls and Latin-1 Supplement | U+0080 – U+00FF |
| Latin Extended-A | U+0100 – U+017F |
| Latin Extended-B | U+0180 – U+024F |

³ Elemental symbols within an agreed set, representing a character for the purposes of writing.

| | |
|--|-----------------|
| Latin Extended-C | U+2C60 – U+2C7F |
| IPA Extensions | U+0250 – U+02AF |
| Combining Diacritical Marks | U+0300 – U+036F |
| Latin Extended-D | U+A720 – U+A7FF |
| Combining Diacritical Marks Supplement | U+1DC0 – U+1DFF |
| Latin Extended Additional | U+1E00 – U+1EFF |
| Latin Ligatures | U+FB00 – U+FB0F |
| Full-width Latin Letters | U+FF00 – U+FF5E |

MSR2 excluded the following ranges:

- Latin Extended-D; technical use (phonetic)/obsolete/punctuation
- Latin Ligatures; not PVALID in IDNA 2008
- Full-width Latin letters; not PVALID in IDNA 2008

1.3 Inclusion

The panel will need to develop criteria for inclusion in the Latin script table of code points. Examples of such criteria could be:

- The code point is used to write a language with an Expanded Graded Intergenerational Disruption Scale (EGIDS) score between 1 and 4.
- The code point is used to write a language with an EGIDS score of 5 or above, but the language is in **modern use**:
 - Current newspapers use Latin script to write the language.
 - The language is written in the Latin script and spoken by a large number (to be defined) of speakers.

Even if a code point falls under a criterion, there could be a reason (to be defined) why it is not possible to include it in the table.

EGIDS defines these levels⁴:

| Level | Label | Description |
|-------|---------------|---|
| 0 | International | The language is widely used between nations in trade, knowledge exchange, and international policy. |
| 1 | National | The language is used in education, work, mass media, and government at the national level. |

⁴ See www.ethnologue.com/about/language-status.

| Level | Label | Description |
|--------------|---------------------|--|
| 2 | Provincial | The language is used in education, work, mass media, and government within major administrative subdivisions of a nation. |
| 3 | Wider Communication | The language is used in work and mass media without official status to transcend language differences across a region. |
| 4 | Educational | The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education. |
| 5 | Developing | The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable. |
| 6a | Vigorous | The language is used for face-to-face communication by all generations and the situation is sustainable. |
| 6b | Threatened | The language is used for face-to-face communication within all generations, but it is losing users. |
| 7 | Shifting | The child-bearing generation can use the language among themselves, but it is not being transmitted to children. |
| 8a | Moribund | The only remaining active users of the language are members of the grandparent generation and older. |
| 8b | Nearly Extinct | The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language. |
| 9 | Dormant | The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency. |

| Level | Label | Description |
|-------|---------|--|
| 10 | Extinct | The language is no longer used and no one retains a sense of ethnic identity associated with the language. |

1.4 Exclusions

Code points must not be punctuation, or solely for historical, religious text or other specialist use.

To determine whether a code point is in a language in modern use, websites such as Ethnologue including EGIDS, Omniglot, ScriptSource and Unicode (especially the Common Locale Data Repository) and Wikipedia will be used. Other major criteria include the number of speakers and whether there exists a modern literature or newspapers in the language.

Certain characters are only used for historical purposes. For example, some consonants in Irish Gaelic were formerly written with a dot above them, e.g., *ḃ, ċ, ḋ, ḟ, ġ, ṁ, ṗ, ṡ* and *ṫ*; now they are written: *bh, ch*, etc.

It is possible that there may not be precomposed forms in Unicode 8.0 for all letters in languages in modern use or even letters that cannot be represented in Unicode 8/0.

The Latin script is often used to Romanize other languages. For example, in the Hepburn Romanization of Japanese, 東京 would be written as *Tōkyō*. Romanization may require the use of unusual diacritics, for example, a dot under a consonant (for example, *ḍ*) may represent a retroflex sound, as in Indian languages. Many unofficial Romanizations also exist such as Arabic chat:
ana raye7 el gam3a el sa3a 3 el 3asr.

Use in a Romanization is not a criteria for inclusion, unless the Romanization counts as a language in modern use, with newspapers, literature, etc. written in it. The Pinyin Romanization of Mandarin is a borderline case in this respect, with further work required on whether its code points, for example *ǔ*, should be included or not.

1.5 Foundation documents and RFCs

Terminology Used in Internationalization in the IETF (RFC 6365) is used for definitions.

The normative statement of the protocol-valid code points is given in RFC 5892 with a corresponding reference table in the IANA Protocol Registry.

Code points in IDNs in Latin script must be PVALID in the IDNA 2008 protocol and CONTEXT O/J.

The work of the panel is to be based on Considerations in the use of the Latin script in variant internationalized top-level domains: *Final report of the ICANN VIP Study Group for the Latin script* and especially on *Procedure to develop and maintain Label Generation Rules for the Root Zone in respect of IDNA labels*.

1.6 Principal languages using the script

Major world languages using the Latin script include:

- Europe: Many Romance, Germanic and Slavonic, and some other languages including Spanish, French, Italian, Portuguese, English, German, Dutch, Swedish, Danish, Norwegian, Polish, Czech, Croatian, Finnish and Hungarian.
- America: Many European languages plus indigenous languages including Guaraní, Cubeo, Q'eqchi', Shavante, Ixil, Zapotec, Atikamekw, etc.
- Eskimo-Aleut: Inuit and Yupic languages, and Aleut.
- Africa: Many European languages plus indigenous languages including Swahili, Hausa and Yoruba.
- Central Asia: Azeri, Turkish, Turkmen, Uzbek, etc.
- Australasia and South-East Asia: Many European languages plus Pitjantjatjara, Maori, Indonesian, Bahasa Malaysia, Tagalog, Vietnamese, Polynesian languages, etc.

See Appendix A for a longer but probably non-exhaustive list.

Europe

- The Latin script is the writing system in widest use in Europe. Cyrillic is used by several countries, for example Bulgaria and Serbia (the latter also widely uses Latin script unofficially), and the Greek alphabet is used in Greece.
- Many languages have modified letters by adding diacritics, for example, *q* in Polish or created digraphs, for example, *œ* in French or new letters, for example *þ* (thorn) in Icelandic.

Americas

- Thousands of languages were spoken before contact with Europeans.
- Many are now critically endangered, with only about ten with an EGIDS score between 1 and 4, but some have been given official status, for example, Guaraní and Aymara.
- Several hundred indigenous languages belonging to many language families are or were spoken in North America.
- Creoles, stable natural languages developed from pidgins (simplified languages or mixture of languages used by non-native speakers) are in use, for example, in the Caribbean and South America.
- In Mexico and Central America, Mayan languages are spoken by some six million people. Yucatec Maya alone has about 800,000 speakers.
- In South America about 350 languages, belonging to, for example, the Tupian, Cariban and Macro-Jê language families are spoken.

- The Latin script is used to write most American indigenous languages and creoles. Syllabics (see also the next section) is used to write some Canadian languages. The Maya script was used to write some Mayan languages.

Eskimo-Aleut

- Eskimo languages split into Inuit languages written in Latin and Inuktitut Syllabics and Yupic languages written in the Latin and Cyrillic scripts. Kalaallisut, spoken in Greenland, is an EGIDS level 1 language.
- Aleut is spoken in Alaska. It is an EGIDS level 7 language, using, for example, ǵ circumflex and x circumflex (which has no precomposed form in Unicode 8.0).

Africa

- Today, the Latin script is the writing system in widest use in Africa.
- It is estimated that over 500 out of the 2000 languages spoken in Africa today have orthographies (Bendor-Samuel 1996: p.689), with the vast majority being Latin script-based.
- The Latin script has been significantly extended or modified to represent African languages:
 - Frequently, supra-segmental features such as tone were encoded using super- and subscripted graph(eme)s, such as accent marks.
 - Next to entirely new letters, di-, tri- and quadrigraphs, for example, are often-much used to represent single phonological units.
- A number of code-points are already excluded by the "letter principle" in the MSR, as well as IDNA 2008.

Central Asia

- The languages of the majority of the inhabitants are Turkic: Azeri, Tatar, Turkish, Turkmen, Uzbek, etc.
- Some languages in the area are sometimes and others exclusively written in the Cyrillic or Arabic scripts.
- Some diacritics are used, for example, ü and ş in Azeri, Turkish and Turkmen, and some additional letters are used, for example, ə (schwa) in Azeri.

South East Asia and Australasia

This area contains Polynesian, Australian, Austronesian and Papuan languages.

- Major Polynesian languages include Hawaiian, Maori, Samoan, Tahitian and Tongan. Long vowels may be indicated by macrons, for example, ō.
- There are fewer than 150 Australian languages in modern use. Some use digraphs, and some diacritics, for example, ṛ in Pitjantjatjara.
- There are over 1,000 Austronesian languages, including Bahasa Malaysia, Indonesian, Formosan languages and Tagalog. Most Austronesian languages now use the Latin script, but there is some use of the Arabic script, for example, Jawi.

- Some Austronesian languages are spoken in New Guinea. Most of the over 1,000 languages spoken there are Papuan languages with Latin-based writing systems.

1.6 Related Scripts

As mentioned above, the Latin and Cyrillic scripts developed from the Greek script and share several letters. The Greek, Arabic and Hebrew scripts developed from the Phoenician alphabet, but the relationship is so distant that there is little visual similarity among most related letters among them. The Armenian script may be modelled on the Greek script and a small number of letters are shared.

The Fraktur and Irish Gaelic writing styles of the Latin script are so different that Unicode considers them different scripts.

Cúairt bé mórriac le dlúctiáð
 fíoréinn trí hata mo dea-riocáin
 67. jkqvwxy 7 z & 1234567890:
Ceanannas an cló a úsáidtear anseo.

sample of Irish Gaelic by Arthur Baker

2. Proposed Initial Composition of the Panel

The role of the LGP is to establish the repertoire and Label Generation Rules for top level internationalized domain names in Latin script.

2.1 Panel Chairs and Members (with Expertise)

The current working group includes the following members in alphabetical order:

| No. | Name | Position | Organization | Country | Language Expertise |
|-----|--------------------|----------|---|--------------|---|
| 1 | Tunde Adegbola | Member | African Languages Technology Initiative | Nigeria | |
| 2 | Sarat Assirou | Member | Institute COADY | Ivory Coast | Dioula, Baoulé Bété, Ebríé |
| 3 | Dwayne Bailey | Member | Translate.org.za | South Africa | Afrikaans, Northern Sotho, Venda, Tswana and Southern Sotho |
| 4 | Ahmed Bakht Masood | Member | Pakistan Telecom Authority | Pakistan | Urdu, English |

| | | | | | |
|----|-------------------------------|----------|--|--------------|--|
| 5 | Matthias Brenzinger | Member | University of Cape Town | South Africa | |
| 6 | Eric Brunner-Williams | Co-Chair | CORE | US | English |
| 7 | Chris Dillon | Co-Chair | University College London | UK | English, German, Spanish |
| 8 | Tarkan Doruk | Member | Sanofi | UAE | Turkish |
| 9 | Yashar Hajiyev | Member | Information Policy Analytical Center | Azerbaijan | Azerbaijani, English |
| 10 | Hazem Hezzah | Member | League of Arab States | Egypt | Use of Latin script for Arabic chat language, German |
| 11 | Paul Hoffman | Member | ICANN | US | English |
| 12 | Tarik Merghani | Member | AfTLD | Sudan | |
| 13 | Meikal Mumin | Member | University of Cologne | Germany | German, English, use of Latin script for African languages |
| 14 | Danko Jevtovic | Member | Fondacija | Serbia | Serbian, English |
| 15 | Ngo Thanh Nhan | Member | New York University | US | Vietnamese |
| 16 | Daniel Omondi | Member | Internet Society | Kenya | |
| 17 | Oscar Gabriel Ledesma Piñeiro | Member | Alfa-REDI | Argentina | Spanish, English |
| 18 | Gideon Kiprono Rop | Member | DotConnectAfrica | Kenya | |
| 19 | Jean-Jacques Subrenat | Member | NCUC; Individual Users; NMI/CC; ICG | France | French, English |
| 20 | Mirjana Tasić | Member | National Internet Domain Names of Serbia (RNIDS) | Serbia | Serbian, English |
| 21 | Aysegul Tekce | Member | ICANN | Turkey | Turkish |
| 22 | Bonface Witaba | Member | Global Knowledge Partnership Foundation | Kenya | Swahili |
| 23 | Jiankang Yao | Member | Computer Network Information | China | Mandarin Chinese, Pinyin and English |

| | | | | | |
|--|--|--|--------------------|--|--|
| | | | Center (CNIC, CAS) | | |
|--|--|--|--------------------|--|--|

Relevant expertise

Name: Chris Dillon

Role: Generation Panel Co-Chair, Academia (linguistic)

Designation: Research Associate, University College London

Relevant experience:

- 2012-present: Member of the VIP Chinese Generation Panel (see <https://community.icann.org/display/croscomlgrprocedure/Chinese+Script+GP>).
- 2016-: Member of IDN Implementation Guidelines Working Group
- 2016- Member of Next-Generation RDS Working Group
- 2014-2015 Formerly Co-Chair of the GNSO Translation & Transliteration of Contact Information Policy Development Project Working Group (see <https://community.icann.org/display/tatcipdp>).
- 2011-2014 Member of the JIG [ccNSO/GNSO Joint IDN Working Group] (see <http://ccnso.icann.org/workinggroups/jiwg.htm>).
- 08/2012–12/2012 Project 2.1 (Root IDN Table Process) (see www.icann.org/en/news/announcements/announcement-3-21-mar13-en.htm).
- Formerly member of the Variant Issues Project Chinese Case Study (see <https://community.icann.org/display/VIP>).
- 2010-2012 Project Manager of the String Similarity Evaluation Panel during the first round of ICANN's New gTLD Program

Name: Sarat Assirou

Role: Linguistic Expert / Community Representative

Designation: Instructor, Institut de Linguistique Appliquée, Université Félix Houphouet Boigny de Cocody, Abidjan, Ivory Coast

Relevant experience

- Linguist, specialist in functional alphabetization
- Consultant on alphabetization
- Lecturer in the department of language sciences at the Université Félix Houphouet Boigny de Cocody

Seminars and experiences on alphabetization:

- October 2007: Participation in the editing seminar to set up the Institutions for Training and Women's Education (Institutions de Formation et d'Education Féminine - IFEF) in Cote d'Ivoire.
- July-August 2010: Realization (in association with Dr Kalilou TERA) of the diagnostic study on alphabetization in Côte d' Ivoire, sponsored by the National Ministry of Education (MEN) with financial support from UNICEF.
- October 2011: Seminar on the validation of the diagnostic study on alphabetization in Abidjan, Côte d'Ivoire (AIBEF).
- Since January 2012: TV presenter under the heading "PARLONS NOS LANGUES" ("Let's speak our Languages) in the broadcast "LES TRESORS DU MONDE" ("Treasures of the World") on channel TV2 on Radio Télévision Ivoirienne (RTI).
- November 2014: Training officers for the direction of Alphabetization and Informal Education (Direction de l'Alphabétisation et de l'Education Non Formelle - DAENF) in methods and techniques for the alphabetization of national languages

Name: Ahmed Bakhat Masood

Role: Regulator, DNS, Arabic Generation Panel, Security
Designation: Deputy Director (ICT/Network)/ Pakistan Telecom Authority
Relevant experience

- 2013 to present: Member of Task Force on Arabic IDN (TF-AIDN)
- 2014- to present: Member of Program Committee Middle East DNS Forum (<http://www.mednsf.org/en/program-committee/>)

1998 to present: Pakistan Telecom Authority (PTA)

- Initiation of different ICT projects for community development like IXP for Pakistan
- Coordination for IPv6 Task Force for Pakistan Network Management, Network Security including DNSSEC and Network forensic
- Coordination with APNIC, SANOG, ICANN and academia for trainings on modern technologies like IPV6, DNSSEC, IRM
- Network and Security management
- Implementation of ISO 27001 standards in PTA

Name: Hazem Hezzah

Role: Arabic Generation Panel member, National and regional policy makers

Designation: IT Expert for ICT Development / League of Arab States

Relevant experience:

- 2013-present: Member of the Task Force for Arabic Script IDNs (TF-AIDN)
- 2012- present: Member of the Multistakeholder advisory group and preparation team for the Arab Internet Governance Forum.
- 2012-present: Participated in preparation, evaluation and contracting for the (.arab) gTLDs, and currently preparing policies for the new gTLD.
- 1991-2011: Performed various IT related roles as support, consultant and technical project manager.

Name: Meikal Mumin

Role: Linguist

- Designation: Institute for African Studies and Egyptology, University of Cologne
- Relevant experience:
- Member of Arabic Generation Panel
- Member of Task Force on Arabic Script IDNs (TF-AIDN)
- Expertise in Roman/Latin script usage for a number of African languages, as well as a general overview of further scripts used in Africa. Active knowledge of German, English, Italian, and French, and familiarity with the writing traditions of those languages and further languages of Modern Europe. Also some familiarity with languages of the Middle East including Arabic and Persian.

Name: Jean-Jacques Subrenat

Role: Policy Expert

Relevant experience

Currently:

- Member of the NTIA IANA Functions' Stewardship Transition Coordination Group (ICG)
- Member of the NETMundial Coordination Council
- President of the Steering Committee, IndividualUsers.org (elected in October 2015)

Member of the ICANN Board of Directors 2007-10 during which:

- Member of President's Strategy Committee (where he was a co-author of the "Implementation Plan for Improving Institutional Confidence")

- Structural Improvements Committee; Public Participation Committee (as its first Chair)
- Member of Board Working Groups: ALAC Review, Board Review, ccNSO Review (as its Chair)

Name: Mirjana Tasić

Role: Registry / DNS/Unicode Expert / Linguist

Designation: Executive Advisor, RNIDS (Register of National Internet Domain Names of Serbia)

Relevant experience

- 08/2012–12/2012 ICANN IDN variant TLD Program: Project (P2.1) - Procedure to Develop and Maintain the Label Generation Rules for the DNS Root Zone in Respect of IDNA Labels ICANN volunteer
- 03/2009 – present: Executive Advisor at RNIDS (Register of National Internet Domain Names of Serbia). Introduction and implementation of IDN ccTLD Fast Track Process for ccTLD <cpb><xn—90a3ac>: string evaluation, domain delegation, sunrise and open registration.
- 07/2006–03/2009 Acting Director of RNIDS (volunteer work) Realization of organizational, political and financial prerequisites for the establishment of RNIDS: RNIDS registration; provision of legal framework for RNIDS operation; organization and establishment of RNIDS office; preparation and implementation of .rs landrush procedures; organization and implementation of the transition process from .yu to .rs domain.
- 04/2006–07/2006 Founder of RNIDS (volunteer work). Organized the RNIDS founding assembly meeting on July 7, 2006.
- 04/1994–09/2008 YU TLD (YU Top Level Domain) Administrator (volunteer work). Managed operation of .yu DNS; Maintained database of .yu domains.
- 1992–1994 Chairwoman, Technical Committee, Academic Network of Yugoslavia. Actively participated in the introduction of internet in Serbia. (volunteer work)
- 1991–10/2010 Administrator of Class B IP address (147.91) assigned to the University of Belgrade, Serbia. (volunteer work)

2.2 Panel Diversity

As the Latin script is used by several hundred languages (see the appendix), it is not possible to have representation from experts of all of them. The approach taken, therefore, is to have experts covering areas of languages, for example, African languages using the Latin script.

? *National and regional policy makers*

Some members of the panel are well versed in ICANN policy, others in national and regional policy.

? *Technical community (general and DNS)*

Although the panel lacks technical expertise, XML training and the LGR Toolset (which automatically generates XML code point by code point) are available.

? *Security and law enforcement*

The panel has little expertise in this area. It is possible that some code points that otherwise would have been included will need to be excluded for security reasons such as lack of compatibility with IDNA or visual similarity. The panel will bear in mind

the sentence in the Procedure to develop and maintain Label Generation Rules for the Root Zone in respect of IDNA labels: “Finally, in investigating the possible variant rules, Generation Panels should ignore cases where the relation is based exclusively on aspects of visual similarity.”

❌ *Academia (technical and linguistic)*

The panel has good coverage of European languages (Romance, Germanic and Slavonic), some coverage of North American indigenous languages, some coverage of African languages, but only weak coverage of South East Asian and especially Central Asian languages and again weak coverage of Australasian languages.

✅ *Community-based organizations*

Several members of the panel work for community organizations.

✅ *Local language computing using Unicode and specifically IDNs*

Several of the linguists have a good knowledge of local language computing, Unicode, IDNA and ICANN’s Variant Issues Project.

2.3 Relationship with Past Work or Working Groups

Until the advent of IDNs in 2003, the “LDH set” – Latin letters “a” to “z” in both upper and lower case, the digits “0” to “9” and the hyphen was used for the registration of names in the DNS.

IDNA (Internationalized Domain Names in Applications) is the protocol used for implementing IDNs. The latest version is 2008, but changes from the 2003 version are likely to break the Longevity Principle in the *Procedure to develop and maintain Label Generation Rules for the Root Zone in respect of IDNA labels*.

ICANN’s Variant Issues Project Study Group for the Latin Script produced *Considerations in the use of the Latin script in variant internationalized top-level domains* in 2011.

3. Work Plan

3.1 Suggested Timeline with Significant Milestones

The Generation Panel intends to divide the work on the LGR for the Root Zone into four stages:

1. Finalization of Code Points
2. Finalization of Variants (if any)
3. Finalization of Whole Label Rules
4. Finalization of LGR Documents for Latin Script and Submission to ICANN

At all stages there will be consultation with the Integration Panel, the Generation Panels of Related Scripts, and the public via periodic public comments.

1. Finalization of Code Points

This stage involves the listing of code points from the parts of Unicode listed in section 1.1 above. Rows in the list will be coloured red until they are attested as in languages in modern use, at which point the colouring will be changed to black. This situation for attested code points will be represented in an XML file. For the non-exhaustive list of languages using the Latin script that is to be used, see the appendix.

2. Finalization of Variants (if any)

The LGP will decide if it is necessary to declare in-script and/or cross-script variants. In the event of a declaration of either sort of variant, an exhaustive list will be made. This situation will be represented in an XML file.

3. Finalization of Whole Label Rules

The LGP will check that no problems are caused by any default WLE and then list any Latin script-specific WLEs, if, for example, some code point may only occur in certain positions in a label (for example, German ß would be mid-label or label-final only), or may only occur together with certain other code points or ranges of code points. This situation will be represented in an XML file.

4. Finalization of LGR Documents for Latin Script and Submission to ICANN

The proposal document and XML files will be completed, taking into account public comments and the work of the Generation Panels of related scripts (at least Cyrillic and Greek). It is possible that a delay may be necessary at this stage.

3.2 Proposed schedules of meetings and teleconferences

The schedule below roughly presumes the Arabic Generation Panel's schedule. The AGP's experience is likely to speed up the LGP's work. The Latin script, however, is used by a larger number of languages and consists of a larger number of code points; both factors which will slow down the LGP's work. The schedule presumes about four months on work with variants. If they are not declared, this may decrease to as little as one month. It may be necessary to appoint advisors to fill gaps in the panel's experience. The panel is composed largely of volunteers and not all of them will have time at all stages of the work.

| Task name | By | Status |
|---|--------------|---------------|
| Develop call for participation | Tue 06-23-15 | Done |
| Publicly release call for participation | Fri 07-24-15 | Done |
| Meeting | Tue 9-22-15 | Done |
| Face-to-face meeting (Dublin) | Sun 10-18-15 | Done |
| Meeting on character set | Tue 11-10-15 | Done |

| | | |
|--|--------------|-------------|
| Invitation to experts to ensure diversity | Fri 11-20-15 | In progress |
| Meeting on character set | Tue 11-24-15 | Done |
| Meeting on character set | Tue 12-08-15 | Done |
| Meeting on panel-formation proposal | Tue 01-05-16 | Done |
| Meeting on panel-formation proposal | Tue 01-26-16 | Done |
| Meeting on panel-formation proposal | Tue 02-09-16 | Done |
| Face-to-face meeting (Marrakech) | Sun 03-06-16 | Done |
| Meeting on character set | Tue 03-22-16 | |
| Meeting on character set | Tue 04-12-16 | |
| Meeting on general principles for inclusion | Tue 04-26-16 | |
| Meeting on general principles for exclusion | Tue 05-10-16 | |
| Meeting on general principles for deferral | Tue 05-24-16 | |
| Release of character set for public comment | Tue 06-07-16 | |
| Meeting | Tue 06-21-16 | |
| Face-to-face meeting | Mon 06-27-16 | |
| Meeting on finalization of character set | Tue 07-12-16 | |
| Meeting: Discussion on variants | Tue 07-26-16 | |
| Meeting: Are in-script variants needed? | Tue 08-09-16 | |
| Meeting: Are cross-script variants needed? | Tue 08-30-16 | |
| Meeting | Tue 09-13-16 | |
| Meeting | Tue 09-27-16 | |
| Meeting on finalization of variants | Tue 10-11-16 | |
| Meeting: Release of variants for public comment | Tue 10-25-16 | |
| Face-to-face meeting (Puerto Rico) | Sun 10-29-16 | |
| Incorporation of comments from public and IG | Tue 11-29-16 | |

| | | |
|--|---------------|--|
| Finalization of variants | Tue 12-13-16 | |
| Discussion of Whole Label Rules | Tue 01-10-17 | |
| Documenting Whole Label Rules | Tue 01-24-17 | |
| Meeting | Tue 02-07-17 | |
| Meeting on finalization of Whole Label Rules | Tue 02-21-17 | |
| Release of Whole Label Rules for public comment | Tue 03-07-17 | |
| Face-to-face meeting (Europe) | Sun 03-12-17 | |
| Incorporation of comments from public and IG | Tues 03-21-17 | |
| Finalize document | Tues 04-04-17 | |
| Meeting | Tues 04-18-17 | |
| Finalize LGR XML structure | Tues 05-02-17 | |
| Final edits | Tues 05-16-17 | |
| Submission to ICANN | Tues 05-30-17 | |

4. References

Frakes, J., *et al.*, "Considerations in the use of the Latin script in variant internationalized top-level domains: Final report of the ICANN VIP Study Group for the Latin script". Los Angeles, Calif.: ICANN, October 2011).

<http://archive.icann.org/en/topics/new-gtlds/latin-vip-issues-report-07oct11-en.pdf>

Blanchet, M., *et al.* "Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR". Los Angeles, Calif.: ICANN, April 2015.

<https://community.icann.org/download/attachments/43989034/Guidelines%20for%20LGR.pdf>

"Considerations for Designing a Label Generation Ruleset for the Root Zone". Los Angeles, Calif.: ICANN, April 2015.

<https://community.icann.org/download/attachments/43989034/Considerations%20for%20LGR.pdf>

"Requirements for LGR Proposals". Los Angeles, Calif.: ICANN, April 2015.

<https://community.icann.org/download/attachments/43989034/Requirements%20for%20LGR%20Proposals.pdf>

Integration Panel “Root Zone Label Generation Rules – LGR-1 Overview and Summary”. Los Angeles, Calif.: ICANN, February 2016.

Common Locale Data Repository.

www.unicode.org/cldr/charts/28/summary/root.html

www.ethnologue.com

www.omniglot.com

www.scriptsources.org

https://en.wikipedia.org/wiki/History_of_the_Latin_alphabet

https://en.wikipedia.org/wiki/Latin_script

Maximal Starting Repertoire (MSR2).

<https://www.icann.org/resources/pages/reports-2013-04-03-en>

<https://en.wikipedia.org/wiki/Sütterlin>

https://en.wikipedia.org/wiki/Gaelic_type

Klensin, J., “Internationalized Domain Names in Applications (IDNA): Definitions and Document Framework” = RFC 5890 (2010). <http://tools.ietf.org/html/rfc5890>

Fältström, P., ed., “The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)” = RFC 5892 (2010). <http://tools.ietf.org/html/rfc5892>

Hoffman, P., et al., “Terminology Used in Internationalization in the IETF” (2011). = RFC 6365 <http://tools.ietf.org/html/rfc6365>

Sullivan, A., et al., “Procedure to develop and maintain Label Generation Rules for the Root Zone in respect of IDNA labels” (Marina del Rey, California: ICANN, March 2013). <https://www.icann.org/en/system/files/files/lgr-procedure-20mar13-en.pdf>

Bendor Samuel, J., “African languages” (1996 p.689-691). Oxford University Press

Hartell, R.L., ed., “Alphabet de langues africaines”. UNESCO - Bureau Regional de Dakar, 1993

IDNA 2008. See RFCs 5890, 5891, 5892, 5893 and 5895.

<https://tools.ietf.org/html/rfc5895> , etc.

ISO 15924 “Codes for the representation of names of scripts”.

<http://unicode.org/iso15924/iso15924-codes.html>