Proposal for Generation Panel

for Latin Script Label Generation Ruleset for the Root Zone

# General Information

The Latin script[[1]](#footnote-1) is derived from the Greek alphabet[[2]](#footnote-2), 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.

Letter G developed from C and J from I. Letter V and U split and a ligature[[3]](#footnote-3) 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., one of the earliest surviving documents in Latin

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

* 1. Target Script for the Proposed Generation Panel

The role of the Latin Generation Panel is to establish the repertoire and Label Generation Rules for top level internationalized domain names in Latin script. Only code points included in MSR-2 will be considered.

The Latin script has the following specifications:

ISO 15924 code: Latn

ISO 15924 no.: 215

English Name: Latin

The set of code points in the Latin script , as specified by MSR-2, lie in the following Unicode ranges:

|  |  |
| --- | --- |
| Script | Range of Unicode code points |
| 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 |
| IPA Extensions  | U+0250 – U+02AF |
| Combining Diacritical Marks | U+0300 – U+036F |
| Combining Diacritical Marks Supplement |  U+1DC0 – U+1DFF |
| Latin Extended Additional | U+1E00 – U+1EFF |
| Latin Extended-C | U+2C60 – U+2C7F |

MSR-2 excluded the following Latin script UNICODE ranges:

|  |  |
| --- | --- |
| Latin Extended-D | U+A720 – U+A7FF |
| Latin Ligatures  | U+FB00 – U+FB0F |
| Full-width Latin Letters | U+FF00 – U+FF5E |

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

### 1.1.1 Diacritics

**Definition** of **diacritic**. : a mark near or through an orthographic or phonetic character or combination of characters indicating a phonetic value different from that given the unmarked or otherwise marked element.

### Diacritics are 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 , 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 exceptions to the stress rules, for example, dieciséis ‘sixteen’, Cádiz.
* In Vietnamese, the same diacritic would indicate a high rising tone.
* Many diacritics are used for for specialized purposes, like phonetic notation/romanization and therefore may not be part of an actual orthography
* Many languages have modified letters by adding diacritics, for example, ą in Polish (U+0105 LATIN SMALL LETTER A WITH OGONEK) or created digraphs, for example, U+0153 œ LATIN SMALL LIGATURE OE or new letters, for example þ (thorn) in Icelandic.
* A letter with two diacritics, for example, ḉ, 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 first diacritic added (U+00E7 ç + 0301 / COMBINING ACUTE ACCENT).

### 1.1.2 Latin Script as Represented in Unicode

As represented in Unicode, the Latin script has some identical glyphs, for example, 0259 ə (schwa) and 01DD (turned e). The following letters belong to both the Latin and Cyrillic scripts: а, е, ѕ, i, ј, к, м, о, p, с, у, and x (non-exhaustive list). Here only lower case letters are considered, as upper case ones may not be used in IDNs.

It is likely that scripts of African languages, for example, contain letters for which Unicode has no pre-composed forms. It is also possible that combining marks may be required for some languages in widespread modern use.

## 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 and Asia Minor: Azeri, Turkish, Turkmen, Uzbek, etc.
* Oceania and Southeast 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.

## Countries with significant user communities using script

#### **1.3.1 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 graphemes, 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.

#### **1.3.2 Americas**

* Over a thousand languages may have been 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, notably Guaraní, Quechua 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 now used, at least as one option, to write all American indigenous languages and creoles. Syllabics (see also the next section) is used to write some Canadian languages. The Maya script was used historically to write some Mayan languages.

#### **1.3.3 Central Asia and Asia Minor**

* 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. In general, Latin script is not used for the languages centred within the Russian Federation.
* 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.

#### **1.3.4 Eskimo-Aleut[[4]](#footnote-4)**

* Eskimo languages split into Inuit languages written in Latin and Inuktitut Syllabics and Yupik languages written in the Latin and Cyrillic scripts. Kalaallisut, spoken in Greenland, is an EGIDS 1 language.
* Aleut is spoken in Alaska. It is an EGIDS 7 language, using, for example, ĝ circumflex (U+011D) and x̂ circumflex (which has no pre-composed form even in Unicode 8.0).

#### **1.3.5 Europe**

* The Latin script is the script in widest use in Europe. The Cyrillic script is used by several countries, for example Bulgaria and Serbia (the latter also widely uses Latin script unofficially), and the Greek script is used in Greece.

#### **1.3.6 Oceania**

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, 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 for Malay.
* 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.

## Related Scripts

As mentioned above, the Latin and Cyrillic scripts developed from the Greek script and share many 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.

# Proposed Initial Composition of the Panel

### 2.1 Panel Chairs and Members (with Expertise)

The current working group includes first the members of the panel, then the observers, in alphabetical order:

| **No.** | **Name** | **Position** | **Organization** | **Country** | **Language Expertise** |
| --- | --- | --- | --- | --- | --- |
|  | Abdeslam Nasri | Member | ATOS | Algeria |  |
|  | Ahmed Bakhat Masood | Member | Pakistan Telecom Authority | Pakistan | Urdu, English |
|  | Hazem Hezzah | Member | League of Arab States | Egypt | Arabic, German |
|  | Jean-Jacques Subrenat | Member | NCUC; Individual Users; NMI/CC; ICG | France | French, English |
|  | Meikal Mumin | Member | University of Cologne | Germany | German, English, Italian, and French, use of Latin script for African languages |
|  | Michael Bauland | Member | Knipp  Medien und Kommunikation GmbHTechnologiepark | Germany | German, English, Finish |
|  | Mirjana Tasić | Chair | Register of National Internet Domain Names of Serbia (RNIDS) | Serbia | Serbian, English |
|  | Seun Ojedeji | Member | AFRINIC BOD member | Nigeria | English,Yoruba native speaker,Igbo |
|  | Danko Jevtovic | Observer | Register of National Internet Domain Names of Serbia (RNIDS) | Serbia | Serbian, English |
|  | Jiankang Yao | Observer | Computer Network Information Center (CNIC, CAS) | China | Mandarin Chinese, Pinyin and English |
|  | Matthias Brenzinger | Observer | University of Cape Town | South Africa |  |
|  | Tarik Merghani | Observer | AfTLD | Sudan |  |

**Relevant expertise**

| **No.** | **Name** | **Role** | **Designation** | **Relevant experience** |
| --- | --- | --- | --- | --- |
|  | Abdeslam Nasri | ICT Architect/ Arabic Generation Panel | ICT Architect and Project Manager / AtoS | * 2014 to present: Member of the Arabic GP
* 2014 to present: Member of the Task Force on Arabic IDN (TF-AIDN)
* Expertise in various IT domains like software development, Internet development and multi-tiered architectures, Enterprise architecture. PSPO I and TOGAF certification
* Panellist at the Internet Governance Forum
 |
|  | Ahmed Bakhat Masood | Regulator/ DNS/ Arabic Generation Panel/Security | Deputy Director (ICT/Network)/ Pakistan Telecom Authority | * 2013 to present: Member of Task Force on Arabic IDN (TF-AIDN)
* 2014- to present: Member of Program Committee Middle East DNS Forum)
* 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
 |
|  | Hazem Hezzah | Arabic Generation Panel member/ National and regional policy maker | IT Expert for ICT Development / League of Arab States | * 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.
* Languages: English, German, use of Latin script for Arabic chat langauge
 |
|  | Jean-Jacques Subrenat | Policy Expert | Président, IndividualUsers.org (European Individual Users’ Association) | * 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)
 |
|  | Meikal Mumin | Linguist | Institute for African Studies and Egyptology, University of Cologne: | * 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.
 |
|  | Michael Bauland | DNS/Registy/ Registrar / IDN | Knipp GMBH | * Development of IDN table for <بازار‎> (.bazaar)
* Senior Software Engineer
* Senior Research Assistant at Leibniz University of Hanover
 |
|  | Mirjana Tasić | Registry / DNS/Unicode Expert / IDN | Executive Advisor, RNIDS (Register of National Internet Domain Names of Serbia) | * 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
* Introduction and implementation of IDN ccTLD Fast Track Process for ccTLD <срб><xn—90a3ac>: string evaluation, domain delegation, sunrise and open registration.
* 07/2006–03/2009 Acting Director of RNIDS (volunteer work). 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).
* 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)
 |
|  | Seun Ojedeji | DNS/ Policy / Native speaker | Chief Network Engineer at Federal University of Oye-Ekiti | * AFRINIC Member Board of Directors
* Open Source Foundation for Nigeria Vice President
* FOSSFA Council Chair
* AFRNIC Policy Development Working Group co-chair
* Principal Network Engineer
* System analyst/Network engineer
 |
|  | Danko Jevtovic | Observer |  |  |
|  | Jiankang Yao | Observer |  |  |
|  | Matthias Brenzinger | Observer |  |  |
|  | Tarik Merghani | Observer |  |  |

### 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.

Because of the panel’s wide remit, the intention is for it to remain open to new members throughout its work. Those without short CVs and currently marked as observers in this document may easily become members.

#### **2.2.1 Overview - Diversity vs Panel members**

|  |  |
| --- | --- |
| **Diversity** | **Panel members** |
| National and regional policy makers | Abdeslam Nasri* Ahmed Bakhat Masood
* Hazem Hezzah
* Jean-Jacques Subrenat
* Seun Ojedeji
 |
| Technical community (general and DNS) | Abdeslam Nasri* Ahmed Bakhat Masood
* Michael Bauland
* Mirjana Tasić
* Seun Ojedeji
 |
| Security and law enforcement | Ahmed Bakhat Masood |
| Academia (technical and linguistic) | * Meikal Mumin
 |
| Community-based organizations | Jean-Jacques Subrenat* Mirjana Tasić
 |
| Local language computing using Unicode and specifically IDNs | Ahmed Bakhat Masood* Hazem Hezzah
* Meikal Mumin
* Michael Bauland
* Mirjana Tasić
* Seun Ojedeji
 |

Panel members will try to reach out experts who are not formally members of the panel. Experts would be advisors to the panel.

The panel will address ICANN to provide necessary advisors for the areas where is the lack of expertize in the panel.

2.2.2 Overview - Language Classification versus Panel members

Following table was derived from the set of Languages using Latin script with EGIDS 1-4 . Sorting order and population number were calculated for highest level of language classification

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NO | Classification | No of languages | Population | GP members expertise | Comment |
|  | Afro-Asiatic | 4 | 42.103.900 |  |  |
|  | Austronesian | 42 | 289.345.146 |  |  |
|  | Creole | 13 | 12.040.180 |  |  |
|  | Indo-European | 41 | 1.359.804.062 | * Mirjana Tasić
 |  |
|  | Mayan | 5 | 4.211.480 |  |  |
|  | Niger-Congo | 36 | 203.298.610 |  |  |
|  | Nilo-Saharan | 4 | 8.457.500 |  |  |
|  | Not grouped | 16 | 107.336.480 |  | The list of languages could be found in Table xx |
|  | Sino-Tibetan | 3 | 3.696.750 |  |  |
|  | Turkic | 9 | 156.055.170 |  |  |
|  | Uralic | 7 | 16.415.980 |  |  |
|  |  | 180 | 2.202.765.258  |  |  |

### 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.

## 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
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.

#### **3.1.1.** **Finalization of Code Points**

This stage involves the listing of PVALID code points from the parts of Unicode included in MSR-2 as listed in section 1.2 above. Each code point will be evaluated and its attestation status indicated. This situation 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.

#### 3**.1.2. Finalization of Variants (if any)**

The LGP will list in-script and/or cross-script variants. This information will be added to the XML file. It is expected that variants will be blocked. That means that if, for example, labels aaıaa, aaɩaa and aaiaa (where the first contained 0131 LATIN SMALL LETTER DOTLESS I and the second LATIN SMALL LETTER IOTA which were blocked variants of LATIN SMALL LETTER I in the third) were applied for in that order, the first application would block the two subsequent applications.

#### **3.1.3. 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. This would be the case, 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 the XML file.

#### **3.1.4. 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 its work. The schedule presumes about four months on work with variants. 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.

### 3.3 Sources for funding travel and logistics

### 3.4 Need for ICANN provided advisors

## 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](http://www.unicode.org/cldr/charts/28/summary/root.html)

[www.ethnologue.com](http://www.ethnologue.com)

[www.omniglot.com](http://www.omniglot.com)

[www.scriptsource.org](http://www.scriptsource.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/S%C3%BCtterlin)

<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>

1. *Script* is used here to indicate the whole writing system including basic letters, ligatures and diacritics. See also RFC 6365 and ISO 15924. [↑](#footnote-ref-1)
2. *Alphabet* is used to refer to the basic set of letters, as used, for example, in a dictionary. [↑](#footnote-ref-2)
3. In writing and **typography**, a **ligature** occurs where two or more graphemes or letters are joined as a single glyph [↑](#footnote-ref-3)
4. These languages are filed under Asia East in the appendix, as are Ainu and Okinawan. [↑](#footnote-ref-4)